Urban greenspace, bluespace, stress, and depression during pregnancy: San Francisco Bay Area

A Thesis submitted to the faculty of
San Francisco State University
In partial fulfillment of
the requirements for
the Degree

Master of Science

In

Geographic Information Science

by

Ana Carolina Pelegrini Guimaraes

San Francisco, California

May 2021

Certification of Approval

I certify that I have read Urban greenspace, bluespace, stress, and depression during pregnancy: San Francisco Bay Area by Ana Pelegrini Guimaraes, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirement for the degree Master of Science in Geographic Information Science at San Francisco State University.

Jerry Davis, Ph.D.
Professor,
Thesis Committee Chair

Nancy Wilkinson, Ph.D.
Professor

Lara Cushing, Ph.D.
Assistant Professor, Env. Health Sciences UCLA

Anna Studwell, M.A.
Associate Director
Institute for Geographic Information Science, SFSU

Urban greenspace, bluespace, stress, and depression during pregnancy: San Francisco Bay Area

Ana Carolina Pelegrini Guimaraes
San Francisco, California
2021

This research explores the relationship between environmental factors and mental health, more precisely, the connection between vegetation ('greenspace'), water bodies ('bluespace'), and how they relate to perceived stress and depression during pregnancy. Questionnaires about mental health were collected from 824 pregnant women in the greater San Francisco Bay Area of California and analyzed in the context of the proximity of their residences to greenspace and bluespace. I used four different datasets to calculate greenspace, including one vector and three raster sets: GreenInfo network (vector), National Land Cover Database (NLCD), Normalize Vegetation Index (NDVI) MODIS, and NDVI Sentinel. Bluespace was calculated using only the land cover dataset. I measured the distance to the closest open access park, the percentage of green and blue areas within four Euclidean buffers (300m, 500m, 1km, and 2km), and the average NDVI within the same buffers distances. In order to test the hypothesis that blue and green spaces are associated with mental health, I calculated Spearman's correlation coefficients, compared average green and bluespace measures across two sub-groups (low vs. high scores for perceived stress and depression), and the average perceived stress and depression scores across categorical greenspace quartiles. From Spearman's correlation, greenspace was in general negatively correlated with depression (rho between -0.11 and 0.03 depending on the greenspace measure) and stress (between -0.06 and 0.02). Women with low levels of perceived stress and depression had higher average levels of greenspace, where it had a strong association observed with NLCD, MODIS, and Sentinel greenspace measures. For bluespace, women with low levels of perceived stress and depression had higher average levels of bluespace, within all NLCD buffers; however, the majority of the significance of bluespace was with perceived stress outcomes. NDVI and NLCD datasets presented a more significant result than the vector dataset (GreenInfo Network); the difference between raster resolutions did not appear to be significant. In the categorical greenspace quantile analysis, greenspace was more significant for depression than for perceived stress. In conclusion, this study found significant connections between mental health and residential proximity to green and blue spaces. Further research is needed to consider distance to green/bluespace as street network analysis, different approaches to calculate the environmental quality, and analysis including socio-economic factors.

Acknowledgments

There are several people I would like to acknowledge. Foremost, I would like to express my deepest appreciation to my thesis committee Jerry Davis, Nancy Wilkinson, Lara Cushing, and Anna Studwell for their continuous support, patience, motivation, enthusiasm, and immense knowledge. Thank you to the faculty and staff of San Francisco State University's Department of Geography and Environment for providing me the support, materials, and access to knowledge without hesitating.

Additionally, I thank Rachel Morello-Frosch (University of California Berkeley and Principal Investigator of the Chemicals in our Bodies [CIOB] study) as well as Stephanie Eick, Erin DeMicco and the rest of the CIOB study team at the University of California, San Francisco for providing access to study data and feedback on my early analysis.

Support for this work was provided by the JPB Environmental Health Fellows Program of the Harvard T.H. Chan School of Public Health. The CIOB study was supported by grants RD83543301 from the United States Environmental Protection Agency, P01ES022841 from the National Institute of Environmental Health Sciences, and UG3OD023272 and UH3OD023272 from the National Institutes of Health Environmental influences on Child Health Outcomes (ECHO) program. Moreover, thank you to the Institute of GIScience, which allowed me to come across this research and expanded my knowledge.

I also would like to thank my family and friends for believing in my capabilities even during times where I could not. I thank my professors from the Federal University of Vicosa in Brazil for encouraging me to pursue a master's degree abroad. Last but not least, I am appreciative of the 2019 Pease Scholarship that supported my studies.

Table of Contents

List of Tables vii

List of Figures viii

1. Introduction 1

1.1 Research questions and hypothesis 2

1.2 Literature review of approaches used to characterize proximity to green/blue space 3

1.2.1 Land Use and Land Cover 4

1.2.2 Access to parks 4

1.2.3 NDVI – Normalized Difference Vegetation Index 5

1.2.4 Statistical analysis 6

2. Methods 8

2.1 Patient data 8

2.2 Study area 9

2.3 Green and bluespace measurements 10

2.3.1 NLCD – National Land Cover Database 12

2.3.2 GreenInfo Network - Open access to greenspace 14

2.3.3 NDVI – Normalized Difference Vegetation Index 16

2.4 Statistical analysis 21

2.4.1 Descriptive statistics 21

2.4.2 Correlation 21

2.4.3 Average levels of green and bluespaces stratified by perceived stress and depression 21

2.4.4 Perceived stress and depression scores across greenspace quartiles 22

3. Results 23

3.1 Spearman's correlation analysis 24

3.2 Average levels of green and bluespaces stratified by perceived stress and depression 27

3.3 Perceived stress and depression scores across greenspace quartiles 33

4. Discussion 36

4.1 Spearman's correlation discussion analysis 36

4.2 Average levels of green and bluespaces stratified by perceived stress and depression 38

4.3 Perceived stress and depression scores across greenspace quartiles 41

5. Conclusions 43

References 47

List of Tables

Table 1. Number of patients of Perceived Stress and Depression dataset in each County of the study area. 10

Table 2. Dataset descriptive statistics - Mean and range of greenness/blueness samples across perceived stress and depression dataset. 23

Table 3. Low and high values of perceived stress and depression. Values of median, 25th, and 75th percentile of green and bluespace variables within each of those groups. 30

Table 4. Perceived stress and depression scores across quartiles of greenspace 35

List of Figures

Figure 1. Pregnant women dataset - illustration of the dataset of perceived stress and depression 9

Figure 2. Study area (Counties): Greater San Francisco Bay Area, CA – USA 9

Figure 3. Schematic of data sources and methods 11

Figure 4. Example of the intersection of the dataset with buffer zones 13

Figure 5. Calculating the percentage of green area in the model builder 14

Figure 7. Example of GreenInfo Network dataset 15

Figure 7. Example of distance to closest greenspace 16

Figure 8. Zonal statistics tool 19

Figure 9. The difference in Sentinel-2 images: original and atmospherically corrected...... 20

Figure 10. Spearman's correlation scatter plots, histogram, and coefficients between health data and 300m buffer size green and blue space variables. 25

Figure 11. Spearman's correlation scatter plots, histogram, and coefficients between health data and 500m buffer size green and blue space variables. 26

Figure 12. Spearman's correlation scatter plots, histogram, and coefficients between health data and 1000m buffer size green and blue space variables. 26

Figure 13. Spearman's correlation scatter plots, histogram, and coefficients between health data and 2000m buffer size green and blue space variables. 27

Figure 14. Box plot of low and high perceived stress for each environmental variable....... 31

Figure 15. Box plot of low and high depression for each environmental variable........ 32

1. Introduction

Contact with nature is fundamental to humanity. Recognizing where 'greenspace or bluespace' are located is one way to help identify areas where humans can come into contact with nature. Many studies have associated the contact with nature with individuals' general health (de Vries et al., 2003; Kondo et al., 2018; Reid et al., 2018), well-being (Carrus et al., 2015; White et al., 2013) physical health (Oreskovic et al., 2015; Triguero-Mas et al., 2015), birth outcomes (Casey et al., 2016; Cusack, 2017; Cusack et al., 2017; Laurent et al., 2013; Yin, 2019), mortality (Gascon et al., 2016; Lachowycz & Jones, 2014; Villeneuve et al., 2012), chronic disease (Wolfe et al., 2014), among other health-related subjects. Studies have found that living surrounded by natural environments has a positive impact on mental health (Alcock et al., 2014; Beyer et al., 2014; Brooks et al., 2017; Gascon et al., 2015; Triguero-Mas et al., 2015; Wheeler et al., 2012; White et al., 2013). Gascon et al. (2015) suggest that the simple fact of observing natural outdoor spaces can be beneficial to the individual's well-being and consequently life expectancy and health (restoration theory).

The term 'greenspace' or 'greenspace' is used to refer to vegetation such as trees, grass, forests, parks, etc., and 'bluespace' or 'bluespace' to visible surface water such as lakes, rivers, oceans, among others (Gascon et al., 2015). There are a few research papers that investigate greenness with pregnancy outcomes and a number of them reported a beneficial effect on birth weight associated with exposure to greenspaces (Dadvand et al., 2012; Laurent et al., 2013; Nieuwenhuijsen et al., 2014). The influence of bluespace is not as strong as greenspace (Gascon et al., 2016; Triguero-Mas et al., 2015; Völker & Kistemann, 2015), however, there are some studies that show it to be beneficial to people's health (Glazer et al., 2018; White et al., 2013).

Kondo et al. (2018) and Yin (2019) consider that the lack of consistency in greenness approaches could be a significant factor in different research conclusions. According to Yin (2019), the most widely-used greenness measure in other studies was the mean of the Normalized Difference Vegetation Index (NDVI) within a buffer distance of the patient's home. In order to investigate if higher resolution images will result in a stronger relationship with health data, Reid et al. (2018) urge more research using images with different resolutions associated with NDVI. Gascon et al. (2015) in their systematic review called for more research studies on which green or bluespace indicators are more significant in relation to mental health. Kondo et al. (2018) highlight the importance of further research involving the relationship between greenspace, stress, and depression.

1.1 Research questions and hypothesis

The purpose of this research is to improve the understanding of different approaches to calculate green and bluespace in the context of perceived stress and depression scores of pregnant women. The main research questions are: How are self-reported stress and depression levels of pregnant women in the San Francisco Bay Area related to their proximity to green and bluespace? Which measures of greenspace are most strongly correlated with stress and depression, and does the spatial resolution of greenspace measures matter?

Based on the literature, I hypothesize that the most effective representation of greenspace in the context of modeling the relationship between stress/depression and greenspace will be the Sentinel-2 NDVI spatial average (finest resolution dataset available to estimate NDVI). Finer resolution of green data might present higher correlations with stress and depression variables. I predict that higher percentages of green and blue areas and the average NDVI will be associated with decreases in perceived stress and depression scores and that the distance to the closest open park will be lower among patients with low scores of perceived stress and depression.

This study will help describe the relationship between stress/depression and green and bluespace, and will be the first study to relate pregnancy, self-rated perceived stress, self-rated depression, and environmental factors in the greater San Francisco Bay Area.

1.2 Literature review of approaches used to characterize proximity to green/blue space

According to Zhan et al. (2020) the most common published methods of calculating greenspace are NDVI (as a mean from circular buffers around each patient's house), land use or cover (as a percentile of greenspace), proximity to greenspaces (as a categorical variable using a cut-off distance), percentile of tree canopy (from aerial imagery), and distance to nearest greenspace (by Euclidean or network methods).

With the aim to capture human contact with the natural environment, it is very common to use the application of buffer sizes to represent the green or bluespace around the subject's address. Zhan et al. (2020) in their systematic review presented a considerable number of studies using 100, 300, 500, and 1000 meter buffer sizes. In another study, Reid et al. (2018) found that larger buffer sizes, such as 1000m and 2000m, have a stronger association with health, perhaps because larger buffer sizes may be more likely to capture the individual's full exposure profile, such as workspace, school, and commuting area.

1.2.1 Land Use and Land Cover

A way to characterize human exposure to green space and blue space is to classify their surrounding residential area by land use or land cover. Michigan State University defines land use as a "series of operations on land, carried out by humans, with the intention to obtain products and/or benefits through using land resources" (Coffey, 2013). Land cover, in turn, can be defined as "the vegetation (natural or planted) or man-made constructions (buildings, etc.) which occur on the earth's surface. Water, ice, bare rock, sand, and similar surfaces also count as land cover" (Coffey, 2013). Note that vegetation and water can be classified as either land use or land cover.

Land cover metrics have been used for studies requiring the classification of green and bluespace. De Vries et al. (2003) explores a land cover database from the Netherlands, which was used for both green and blue areas. Additionally, Wheeler et al. (2012) explored the use of green and bluespaces in relation to well-being, calculating coastal proximity and the percentage of greenspace within a buffer size from the land-use dataset. Other studies used land cover data to calculate the exposure of people's health to greenness (Lachowycz & Jones, 2014) and blueness (Triguero-Mas et al., 2015).

1.2.2 Access to parks

There is a difference between living close to a natural environment and being able to access it. Access to parks has not been well-studied in a health context. Cusack (2017) applied access to parks in a network analysis of the distance to the closest natural environment in Portland, Oregon, and Austin, Texas. Grazuleviciene et al. (2015) estimated the proximity to city parks using Euclidean distance in Kaunas, Lithuania. Both studies categorized distance as a binary variable (e.g. distances smaller than 300m or not) in order to associate it with the health data.

1.2.3 NDVI – Normalized Difference Vegetation Index

Normalized Difference Vegetation Index (NDVI) is a ratio using near-infrared and red spectral bands that measures the density of greenness on a patch of land using sensors from aerial platforms that capture the quantity of radiometric wavelengths. There are two bands used to calculate NDVI: near-infrared (NIR) centered at 832.8nm and 106nm wide and visible red centered at 664.6nm and 31nm wide (Eick et al., 2020; Myneni et al., 1995). Equation 1 illustrates how NDVI is computed. This equation ranges from -1 to 1, where negative values will represent water and snow, positive values close to zero describe soils and urban areas, and high values represent healthy vegetation (high in chlorophyll).

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad \text{Eq. (1)}$$

Where NDVI is Normalized Difference Vegetation Index;

NIR is the reflectance value of the Near Infrared band's wavelength; and

'Red' is the reflectance value at the red band's wavelength.

NDVI is a useful measure for vegetation health and another common metric in studies characterizing human exposure to greenspace. The use of NDVI is predominant in this area, however, the analysis of different resolutions and sources of NDVI exposure have not been well studied (Reid et al., 2018). Only a few studies consider images with different spatial resolutions. Reid et al. (2018) used multiple satellites' images and different resolutions including Landsat 8 (30x30m), MODIS (Moderate Resolution Imaging Spectroradiometer at 250x250m), and AVHRR (1x1km). Comparing greenspaces from NDVI and health data, they concluded that of the data sources used in their research, Landsat 8 (higher spatial resolution) presented a better correlation with self-rated health than MODIS and AVHRR (lower spatial resolution).

There are numerous sources for NDVI with different spatial resolutions. MODIS is a common source in research associated with health (Casey et al., 2016; Cusack, 2017; Reid et al., 2018). MODIS has available NDVI images atmospherically corrected in 250m, 500m, 1km, and 0.05Deg spatial resolutions and 16-day temporal resolutions (NASA, n.d.-b). Sentinel-2 provides multispectral images in resolutions of 10m, 30m, and 60m, where the panchromatic spectrum is a 10m resolution, about every 5 days (*Sentinel-2 - Missions - Sentinel Online*, n.d.) from which NDVI can be calculated. For analysis of NDVI, it is important to use atmospherically corrected images, especially when comparing datasets from different periods (Mather & Koch, 2011). This is one of the finest freely-available spatial resolution sources, although the use of NDVI with Sentinel-2 images was not found in the literature associated with health analysis.

1.2.4 Statistical analysis

In order to validate the connection and compare environmental factors and health data, different statistical analyses can be applied. Reid et al. (2018) aggregated exposure to greenness along with buffer sizes and compared it with health indicators by calculating Pearson's Correlation Coefficient for each individual as derived from the different vegetation datasets. Cusack et al. (2017) followed the same methodology by calculating the correlation between the many different greenspace metrics and birth weight. The majority of researchers have used regression models to identify the significance of the influence of green or bluespace upon health measures (Beil & Hanes, 2013; Dadvand et al., 2012; Lachowycz & Jones, 2014; Rodríguez et al., 2012; Tamosiunas et al., 2014; Triguero-Mas et al., 2015). Some authors use the comparison of the highest and lowest quartiles of greenspace indicators (Cusack, 2017; Cusack et al., 2017; Wilker et al., 2014).

2. Methods

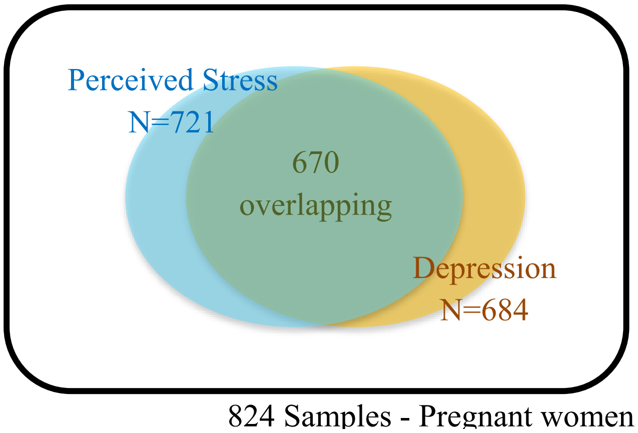
2.1 Patient data

Self-reported data from pregnant women from the Chemicals in Our Bodies (CIOB) cohort was provided by the University of California, San Francisco (Eick et al., 2020). These women lived at 824 addresses and had given birth at two hospitals in San Francisco, California. The first, Moffitt Long/ Mission Bay Hospital, drew women from an economically diverse group; the majority had health insurance. The second, Zuckerberg San Francisco General Hospital, was primarily accessed by low-income women of color who did not have health insurance. The data were collected between 2014 and 2019. The research was approved by the Institutional Review Board of the University of California, San Francisco (Protocol #10–00861) and Berkeley (#2010-05-04). Geocoding of patients' addresses was completed by the CIOB study team using GoogleAPI and the DeGauss geo-coder, with 96% of the geocodes presenting a 75% score match.

A questionnaire was administered in English or Spanish to study participants during the second trimester of pregnancy. The variable for perceived stress was measured using a four-item Perceived Stress Scale (PSS) questionnaire. On a scale of 0 to 16, higher values represent more stress; individuals reporting values of 9 or higher were considered stressed. Depression was calculated using a 10-item Center for Epidemiologic Studies-Depression (CES-D) questionnaire; on a scale of 0 to 30, individuals reporting values of 16 or higher were considered depressed (Eick et al., 2020).

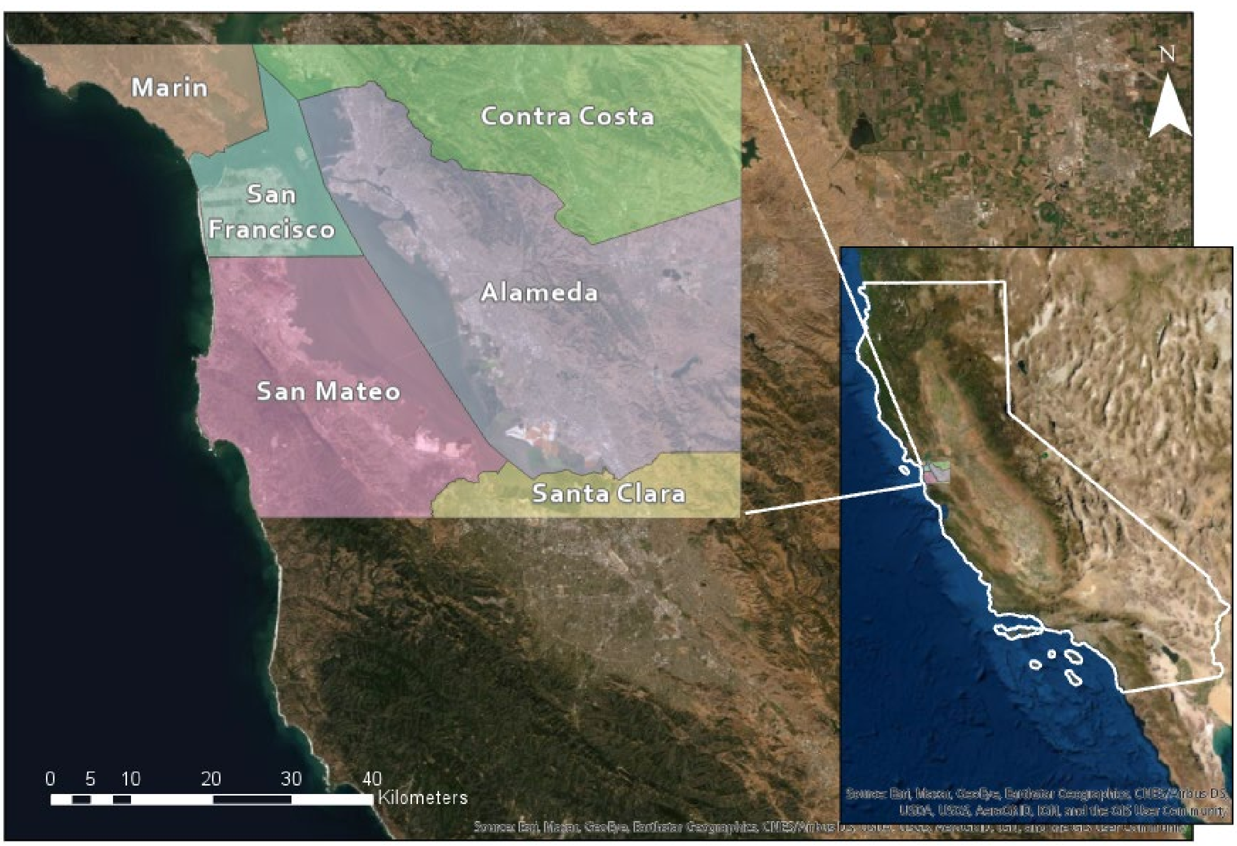
Of the 824 addresses provided, 33 addresses were excluded because they were partially or completely outside the area covered by the imagery used for analysis. Other patients were excluded due to missing data for perceived stress (70 patients) and depression (107 patients). Therefore, the total samples vary for each outcome variable: N=721 for perceived stress and N=684 for depression (figure1).

Figure 1. Pregnant women dataset - illustration of the dataset of perceived stress and depression



2.2 Study area

Figure 2. Study area (Counties): Greater San Francisco Bay Area, CA – USA



This research was done in the San Francisco Bay Area, California, USA. The polygon shown in Figure 2 represents the area where all patients lived during their pregnancy. Patient' addresses are not evenly distributed: the majority lived in San Francisco County as illustrated in Table 1.

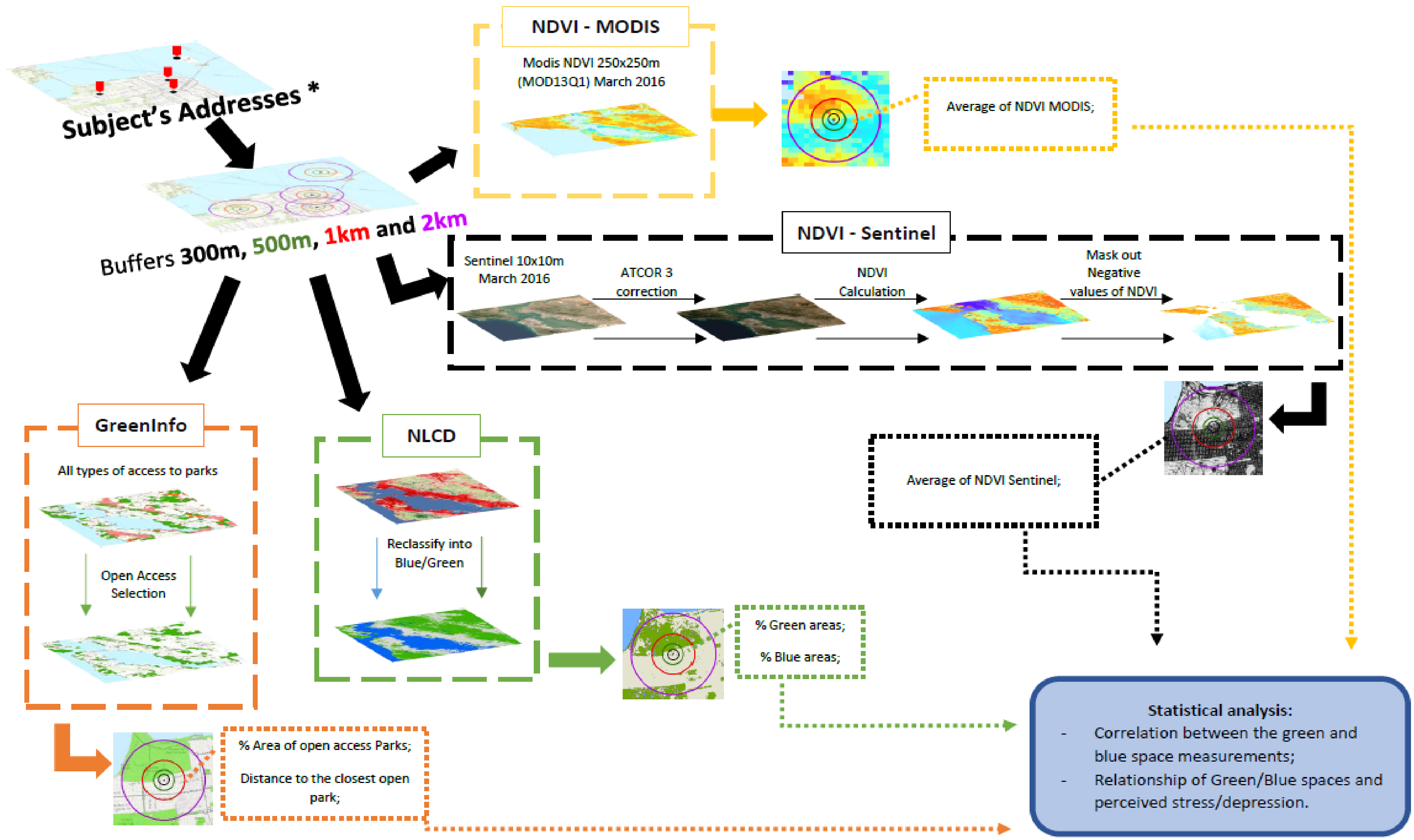
Table 1. Number of patients of Perceived Stress and Depression dataset in each County of the study area.

Counties	Number of Patients from Perceived Stress sample (721 patients)	Number of Patients from Depression sample (684 patients)
San Francisco	618	585
Contra Costa	<= 5	<= 5
Alameda	45	43
San Mateo	46	45
Marin	<= 5	<= 5
Santa Clara	<= 5	<= 5

2.3 Green and bluespace measurements

In order to compare different methods of measuring greenspace, this research used four different data sources: GreenInfo Network, National Land Cover Database (NLCD), and NDVI from MODIS, and NDVI from Sentinel-2 satellite sensors. Figure 3 is a schematic of the method applied in this research.

Figure 3. Schematic of data sources and methods



2.3.1 NLCD – National Land Cover Database

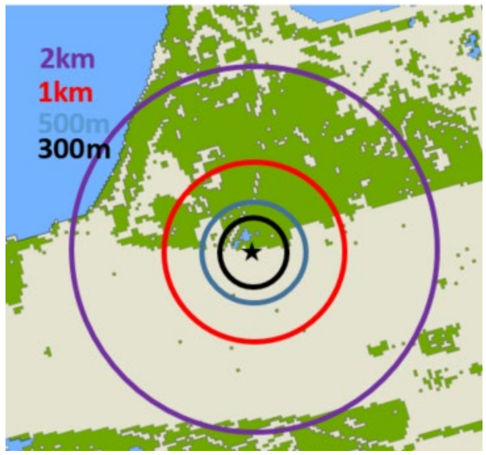
The National Land Cover Database (NLCD) has nationwide data on land cover, using a 15-class legend based on a modified Anderson Level II classification system. The raster dataset is at a 30m resolution based on Landsat imagery, with land cover maps available from 2001 to 2016. This study uses imagery from 2016, the most recent map available, and represents the middle date of the patient sample (*Multi-Resolution Land Characteristics*, n.d.). Considering the spatial resolution of this dataset, the use of NLCD will capture greenspaces/bluespaces that appear in 30-meter resolution as recreational areas, such as the ocean, lakes, national parks, state parks, golf courses, etc. Vegetation or water bodies that do not appear in a 30-meter resolution such as backyards, front yards, neighborhood trees, small gardens, streams, and creeks were not included as green/bluespace.

In order to create meaningful categorical data for this study, I reclassified the NLCD 2016 image into one image of three classes: green space, blue space, and other. Green space included developed open space, deciduous forest, evergreen forest, mixed forest, dwarf scrub, shrub/scrub, grassland/herbaceous, sedge/herbaceous, pasture/hay, cultivated crops, woody wetlands, and emergent herbaceous wetlands (e.g. NLCD codes 21, 41, 42, 43, 51, 52, 71, 72, 81, 82, 90 and 95). Blue space included only open water (NLCD Code 11). While they may be submerged periodically, wetlands can be heavily vegetated thus, for this research, we chose to include them as green space instead of blue.

The spatial reference of NLCD (Albers Conical Equal Area) was different from the projection used for patient data (UTM zone 10). A vector polygon of green and blue spaces based on the NLCD raster was created and transformed to the geographic coordinate system NAD1983 and projection UTM Zone 10 such that the classification of pixels would be in alignment with the patient data and consistent across the study area.

2.3.1.1 Percentage of green and bluespace within the buffer zone

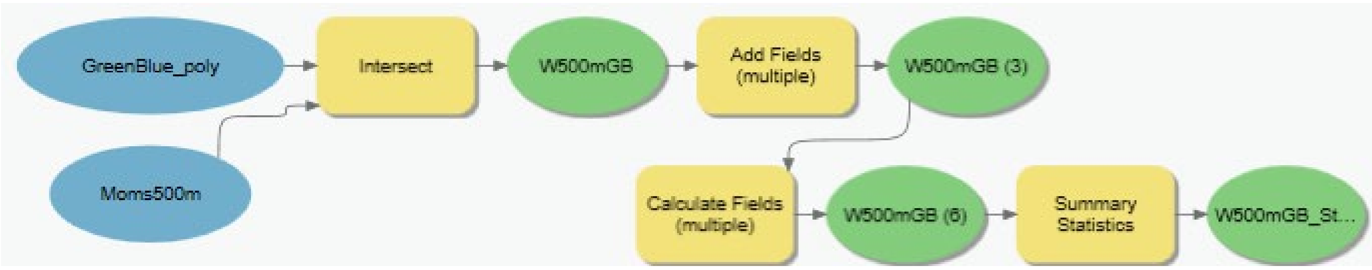
Figure 4. Example of the intersection of the dataset with buffer zones



Fictitious subject's address and its buffer zones intersecting the NLCD reclassified data.
Where the color green, blue and beige represents green space, blue space and other areas, respectively.

A number of studies have calculated the percentage of green and blue spaces within a specific buffer zone (Cusack, 2017; Lachowycz & Jones, 2014). For the 300m, 500m, 1000m, and 2000m buffer distances (Fig. 4) generated around each patient's address, the percentage of area covered by greenspace or bluespace was calculated using a model (Fig. 5; Esri, ArcGIS Pro 2.3.3) that intersected the NLCD-derived greenspace and bluespace polygons with each of these buffers. This procedure enabled us to capture the percentage of green and blue spaces near each patient's home in each of the four buffer sizes, computing a total of eight variables: NLCD Blue 300m, NLCD Blue 500m, NLCD Blue 1km, NLCD Blue 2km, NLCD Green 300m, NLCD Green 500m, NLCD Green 1km and NLCD Green 2km.

Figure 5. Calculating the percentage of green area in the model builder



Model builder (ESRI, ArcGIS Pro 2.3.3) used to calculate the percentage of green and blue spaces on NLCD data for 500m buffer distance.

2.3.2 GreenInfo Network - Open access to greenspace

The GreenInfo Network dataset I utilized, published in June 2019, is an update of the California Protected Areas Database (CPAD) that includes lands that are owned outright and protected for open space purposes. It ranges from the smallest urban parks to the largest wilderness areas and includes national/state/regional parks, forests, preserves, wildlife areas, large and small urban parks that are mainly open space (as opposed to recreational facility structures), land trust preserves and special district open space lands (watershed, recreation, etc.) and other types of open space. This dataset contains more than 97.1% of open access to the public areas; the remaining 2.9% of areas are restricted to non-public and unknown areas (*CPAD Database Manual*, 2019).

Figure 6 illustrates the CPAD dataset within the Google basemap image. Note that the CPAD dataset does not include greenspace such as private lands in general (e.g. yards, private golf courses), public lands not intended for public use (e.g. municipal waste facilities, administrative buildings), military lands, and tribal lands (*CPAD Database Manual*, 2019).

Figure 7. Example of GreenInfo Network dataset

Image of areas that are green environments, but are not included in the data. For example this cemetery in South San Francisco, CA
Font: https://www.calands.org/

CPAD has three levels of data for protected lands: Holdings, Units, and Super Units. The Super Units shapefiles are aggregations of units to create user-focused polygons for each park name. While Super Units are typically used in cartographic representation, their use in mostly recreation-focused applications suggested these areas would be the best available data to use for investigating proximity of patient addresses to the types of open access areas that may have a positive association with health. This source will provide green space for open access parks beyond governmental boundaries (*CPAD Database Manual*, 2019).

Using the GreenInfo Network dataset, I calculated the distance between the closest open access park and each patient's address, as well as the percentage of greenspace (open access park) within the buffer zones of each patient's address (300, 500, 1000, and 2000m).

2.3.2.1 Distance to closest GreenInfo Network greenspace

The nearest open access was computed using the Near tool (*ESRI ArcGIS Pro 2.3.3*), which returned the distance to the closest open-access park in a Euclidean distance (figure 7). This represents access to parks as a continuous variable, where it might be inversely proportional to perceived stress and depression (larger distances represent parks further away from the patient's house).

Figure 7. Example of distance to closest greenspace

Distance to the closest park for a fictitious patient's address (represented as a black star). GreenInfo Network dataset is represented as the green polygons, and the basemap is a topographic representation of the city by ArcGIS Pro 2.3.3.

2.3.2.2 Percentage of GreenInfo Network within buffer

Using the same method as in Figure 5, but based on the GreenInfo Network dataset (Super Units), I calculated the percentage of open access parks within each buffer zone (300, 500, 1000, and 2000m) yielding four variables that indicate the percentage of greenspace near each patient: GreenInfo 300m, GreenInfo 500m, GreenInfo 1km, and GreenInfo 2km.

2.3.3 NDVI – Normalized Difference Vegetation Index

In order to compare the difference in spatial resolutions, two sources of images were considered for the NDVI approach: MODIS and Sentinel-2. MODIS has NDVI available at 250 meters spatial and 16 days of temporal resolution, while Sentinel-2 data provides a finer of 10m for spatial and 5 days for temporal resolution.

2.3.3.1 Selecting the most reliable date to download the Modis and Sentinel images

For this research, the analysis of NDVI was restricted to one date that would represent the greenspace. Notice that the period of pregnancy for all patients varies from 2014 to 2019. March 2016 was selected as the best date to represent the population, due to a few considerations: 2016 is the midpoint of the survey of mothers' pregnancies, and March, as the second rainiest month of 2016, would have seen a maximum amount of greenspace (*Golden Gate Weather - Monthly Rainfall*, n.d.).

The urban environment of San Francisco and the greater bay area have many irrigated areas. Because of that, the majority of patients would not be affected by seasonal greenness, since the irrigation system keeps the vegetation healthy the whole year. However, to make sure I took into account parks that are not irrigated, I considered some other facts. March is the end of the rainy season and the beginning of spring; therefore I assumed that the vegetation would be generally greener. Furthermore, once I compared the images from January and March of 2016, the number of shadows in the March images was minimized, given the fact that the sun declination in the images decreases because it is closer to the summertime where the declination is minimal.

2.3.3.2 MODIS (250 meters spatial resolution)

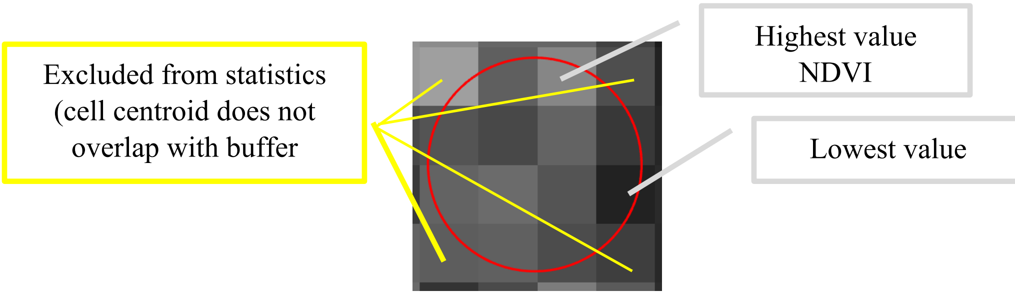
MODIS images were captured from March 5th to March 20th, 2016 (data downloaded on April 17th, 2020 - *MODIS Web*, n.d.). NDVI is derived from images taken over 16-day intervals and at 250 meters spatial resolution and retrieved from daily and atmosphere-corrected images.

The algorithm uses a MODIS-specific compositing method based on product quality assurance metrics to remove low-quality pixels (NASA, n.d.-b) and to derive a measure of NDVI for each pixel.

I calculated the mean NDVI for each buffer using Zonal Statistics for Overlapping Features tool (*ESRI ArcGIS Pro 2.3.3*, pers. communication: Esri Technical Support). This script calculates the statistics of the pixels that have their centroid overlapping the zone, which are the buffer sizes of 300, 500, 1000, and 2000m. The Zonal Statistics for Overlapping Features was a script provided by Esri support, which addresses some of the inconsistencies within overlapping buffers of the ArcGIS Pro 2.3.3 of Zonal Statistics required for this analysis.

The number of pixels associated with the mean NDVI varies according to where each patient's address is located since the buffer might overlap a different number of pixels' centroids. Even when no missing pixels overlap the buffer, the number of pixels might be slightly different for each patient. Furthermore, if a 250x250m MODIS pixel centroid lies outside the buffer, the tool would not count it as part of the equation, resulting in an average over a smaller area than the buffer's size. The reverse is also true: if the pixel has its centroid within the buffer zone, the Zonal Statistics for Overlapping Features equation would result in the calculation of a larger area than the buffer covers. Figure 8 shows that the buffer overlaps with 16 pixels, however, it considers only the 12 pixels with their centroids intersecting the buffer as part of the mean.

Figure 8. Zonal Statistics tool



The pixel within the centroid overlapping the buffer will be included in the statistics. Red circle represents the buffer zone and the pixels in grey scale pixels the NDVI.

The majority of the water pixels appear to be masked out of the raster. Additionally, in most cases, these pixels do not follow the coastal boundary and it masked pixels located on the land as well. For that reason, there are some addresses missing a number of NDVI pixels in their calculation.

Therefore, the use of MODIS images resulted in four variables: MODIS 300m, MODIS 500m, MODIS 1km, and MODIS 2km, which represent the average of MODIS NDVI pixels inside the buffer zone for each patient's address.

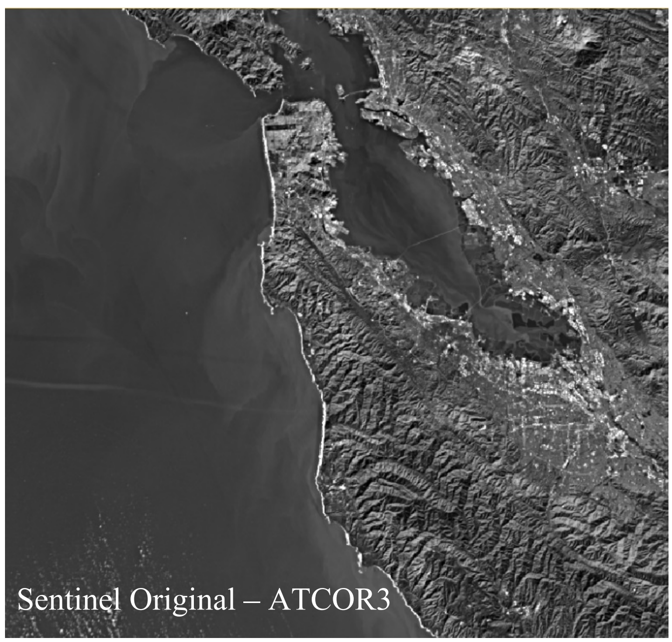
2.3.3.3 Sentinel-2 (10 meters spatial resolution)

Imagery from Sentinel-2 was captured on March 26th, 2016 (data downloaded on May 6th, 2020 - NASA, n.d.). The constellation of two identical satellites of Sentinel-2 is a result of a multispectral instrument that collects 13 spectral bands (four bands at 10m - red and near-infrared are referred here, six bands at 20m and three bands at 60m resolution). The revisit time is about 5 days (10 for each satellite) at the equator and 2-3 days at mid-latitudes (*Sentinel-2 - Missions - Sentinel Online*, n.d.).

I first compared two atmospherically corrected images -- one from January 2019 and another from April 2019 -- to evaluate the impact of the shadows resulting from the sun angle. The number of shadow pixels in downtown San Francisco was substantially smaller in April.

Before the NDVI was calculated, the 2016 Sentinel image was atmospherically corrected using ATCOR-3 Workflow for Imagine (Geosystems) and a 10m DEM (Digital Elevation Model) from USGS (downloaded on March 13th, 2020 - USGS, n.d.) in the software *Erdas Imagine* 2018 (Hexagon Geospatial). Figure 9 illustrates the difference between the original image lacking atmospheric correction and the image atmospherically corrected.

Figure 9. The difference in Sentinel-2 images: original and atmospherically corrected



Raster subtraction of the original Sentinel March 2016 image and the new atmospherically corrected using ATCOR3.

In sequence, the NDVI was calculated using equation 1. Some authors excluded the negative values of the index (Reid et al., 2018) and others maintained it (Triguero-Mas et al. 2015, Casey et al.2016). For this research, the negative values were masked out, since they represent areas of water or shadow (significant in downtown SF). Therefore, 'no data' values were assigned to those negative pixels, and they were not included as part of the mean NDVI. Patients' neighborhoods that are very close to the water would thus have a lot of missing pixels, which might result in a skewed average of the NDVI.

Finally, the mean NDVI was calculated for the buffer sizes using the Zonal Statistics tool as mentioned before for the MODIS dataset (Figure 8). The use of Sentinel-2 images resulted in four variables: Sentinel 300m, Sentinel 500m, Sentinel 1km, and Sentinel 2km, which represent the average of Sentinel-2 NDVI pixels inside the buffer zone surrounding each patient's address.

2.4 Statistical analysis

Descriptive statistics and correlation coefficients were used to assess relationships between measures of green/blue space and perceived stress and depression. Statistics were calculated using the software *R 4.0.2* and *Rstudio 1.3.1056*.

2.4.1 Descriptive statistics

Descriptive statistics (maximum, minimum, quantile, and mean values) were calculated separately for all variables of the perceived stress, depression, green and blue spaces. The Shapiro-Wilk test of normality was used to verify whether all variables were normally distributed; the null hypothesis of normal distribution was rejected for all variables.

2.4.2 Correlation

Due to the non-normality of the data, the Spearman non-parametric rank correlation coefficient was used to estimate associations between the variables.

2.4.3 Average levels of green and bluespaces stratified by perceived stress and depression

Two sub-groups were defined for perceived stress and depression (high and low). The threshold for perceived stress was 9: equal or higher values were considered high stress, values less than 9 were considered low stress. For depression, the threshold was 16: equal or higher values were considered highly depressed, and values less than 16 considered low on the depression scale (Eick et al., 2020).

The mean, 25th, and 75th percentiles of each green and bluespace variable were calculated for both sub-populations: low and high. Since a considerable number of patients did not live close to the water, some of the blue variables had a large number of zeros. In order to describe these variables, the 95th percentile, 75th percentile, and maximum values were calculated.

Lastly, the one-sided non-parametric Wilcoxon Test was used to test the null hypothesis of no difference between the mean rank of low and high groups. Note that the unpaired two-sample Wilcoxon test is an alternative to the unpaired two-sample t-test used when the data is not normally distributed.

2.4.4 Perceived stress and depression scores across greenspace quartiles

Green and bluespace measures were next split into four categories based on quartiles. I then assessed average continuous perceived stress and depression scores across each of the four quartiles. In other words, the greenspace data were arranged into ascending order by quartile groups, and the mean of perceived stress and depression was calculated for each quartile. A one-sided Wilcoxon Test was used to test the null hypothesis of no difference between the first quartile (Q1) and fourth quartile (Q4) of greenspace. I excluded bluespace from this analysis because the majority of patients did not live close to water bodies, resulting in many zero values.

3. Results

The study population included 721 patients self-reporting perceived stress and 684 patients self-reporting perceived depression. Table 2 describes descriptive statistics (mean, minimum, and maximum) of perceived stress, depression, green and bluespace variables for both groups. Perceived stress was measured using scores from 0 to 16, where higher values represented more stress. Depression was measured using scores from 0 to 30, where higher values represented more depressive symptoms. Distance from the patient's house to open access greenspace (GreenInfo Network) was calculated in meters. GreenInfo Network, NLCD green, and bluespaces were calculated as the percentage within each buffer area. Finally, MODIS and Sentinel represent the average NDVI within each buffer distance of the patient's residential address.

Table 2. Dataset descriptive statistics - Mean and range of greenness/blueness samples across perceived stress and depression dataset.

Variable	Perceived Stress (N= 721) Mean (Min – Max)	Depression (N=684) Mean (Min – Max)
Perceived Stress	5.25 (0 - 15)	-
Depression	-	6.96 (0 - 30)
Distance to GreenInfo (m)	251.37 (0.00, 4067.75)	242.74 (0.00, 4067.75)
GreenInfo 300m %	6.05 (0.00, 100.00)	6.05 (0.00, 100.00)
GreenInfo 500m %	7.32 (0.00, 100.00)	7.34 (0.00, 100.00)
GreenInfo 1km %	9.27 (0.00, 88.23)	9.28 (0.00, 88.23)
GreenInfo 2km %	10.74 (0.00, 54.41)	10.76 (0.00, 54.41)
NLCD Green 300m %	7.91 (0.00, 86.75)	7.97 (0.00, 86.75)
NLCD Green 500m %	9.27 (0.00, 82.81)	9.36 (0.00, 82.81)
NLCD Green 1km %	11.04 (0.37, 78.66)	11.07 (0.37, 78.66)
NLCD Green 2km %	12.57 (1.06, 85.36)	12.57 (1.06, 77.06)
Modis 300m (NDVI)	0.28 (0.06, 0.73)	0.28 (0.06, 0.73)
Modis 500m (NDVI)	0.28 (0.07, 0.72)	0.28 (0.07, 0.72)
Modis 1km (NDVI)	0.29 (0.08, 0.70)	0.29 (0.08, 0.70)
Modis 2km (NDVI)	0.30 (0.09, 0.72)	0.30 (0.09, 0.72)
Sentinel 300m (NDVI)	0.25 (0.06, 0.72)	0.25 (0.06, 0.72)

Sentinel 500m (NDVI)	0.26 (0.07, 0.70)	0.26 (0.07, 0.70)
Sentinel 1km (NDVI)	0.27 (0.08, 0.70)	0.27 (0.08, 0.70)
Sentinel 2km (NDVI)	0.28 (0.10, 0.69)	0.28 (0.10, 0.69)
NLCD Blue 300m %	0.36 (0.00, 30.11)	0.35 (0.00, 30.11)
NLCD Blue 500m %	0.88 (0.00, 45.81)	0.86 (0.00, 45.81)
NLCD Blue 1km %	2.70 (0.00, 66.37)	2.63 (0.00, 66.37)
NLCD Blue 2km %	6.37 (0.00, 86.27)	6.30 (0.00, 86.27)

3.1 Spearman's correlation analysis

Figures 10 to 13 show Spearman's correlation matrix of the outcome variables, greenness, and blueness measures for each buffer distance (300m, 500m, 1km, and 2km). Bivariate scatter plots are given below the diagonal (including the correlation ellipses), histograms on the diagonal, and Spearman's correlation coefficient (rho) above the diagonal.

Among the correlations of the environmental variables, the distance to the closest open park presented a high negative correlation using GreenInfo Network (300m and 500m), a moderate negative correlation using NLCD green 300m, and no connection with the other variables. Additionally, NLCD green, MODIS, and Sentinel presented a high positive correlation within each other and it increased with the buffer size. The correlations between MODIS and Sentinel were very high for all buffer zones (rho ranging from 0.92 to 0.99), where the highest correlation was between the 2km buffer area. Nearly every correlation between the greenspaces had a significant p-value. The percentage of bluespace was small and positively correlated with greenspace variables for buffer zones of 300m, 500m, and 1km; however, for the buffers of 2km, bluespace was negatively correlated with variables of greenspace.

Perceived stress and depression outcomes were positively and moderately correlated (rho of 0.56), with statistical significance. The environment variables (green and blue space), stress, and depression presented in general a negative weakly correlated (rho from -0.11 to 0.03). Perceived stress and 2km buffer for NLCD, MODIS, and Sentinel presented the most considerable correlation of negative rho -0.06. On the other hand, depression outcomes had a substantial correlation with NLCD 500m, with a negative rho of -0.11. Notice that NLCD greenspace and Sentinel NDVI exhibited a small but statistically significant inverse correlation with depression at all four buffer distances.

Figure 10. Spearman's correlation scatter plots, histogram, and coefficients between health data and 300m buffer size green and blue space variables.

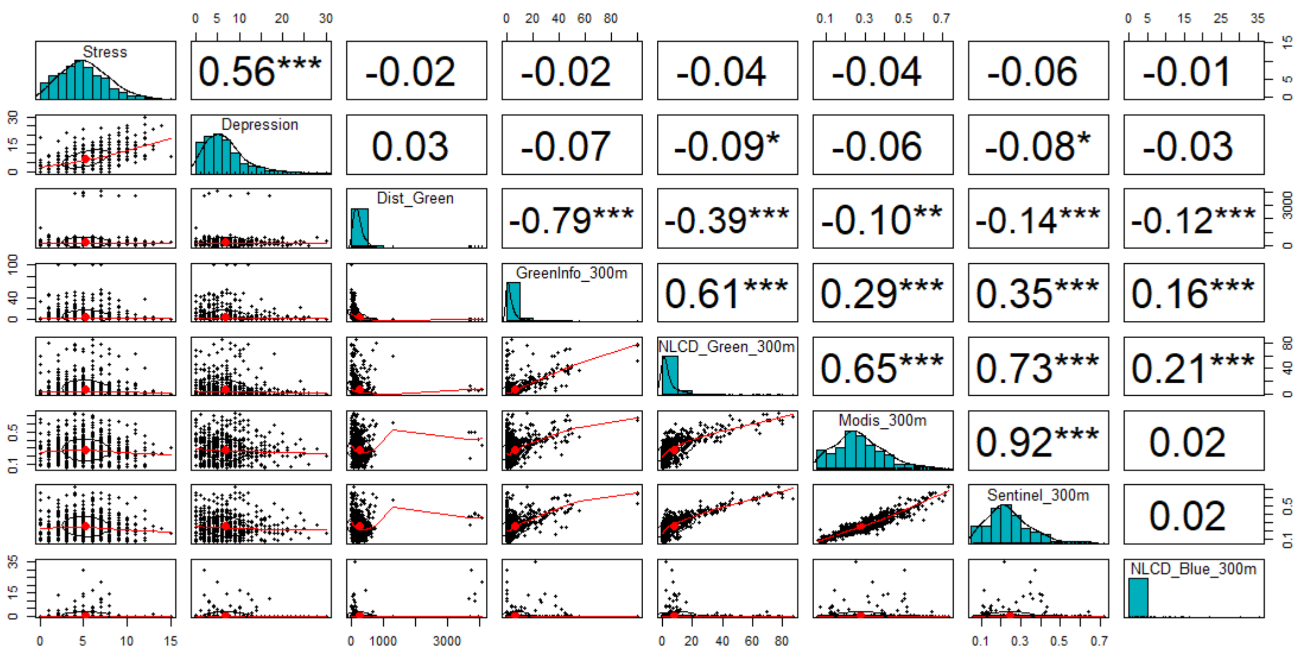


Figure 11. Spearman's correlation scatter plots, histogram, and coefficients between health data and 500m buffer size green and blue space variables.

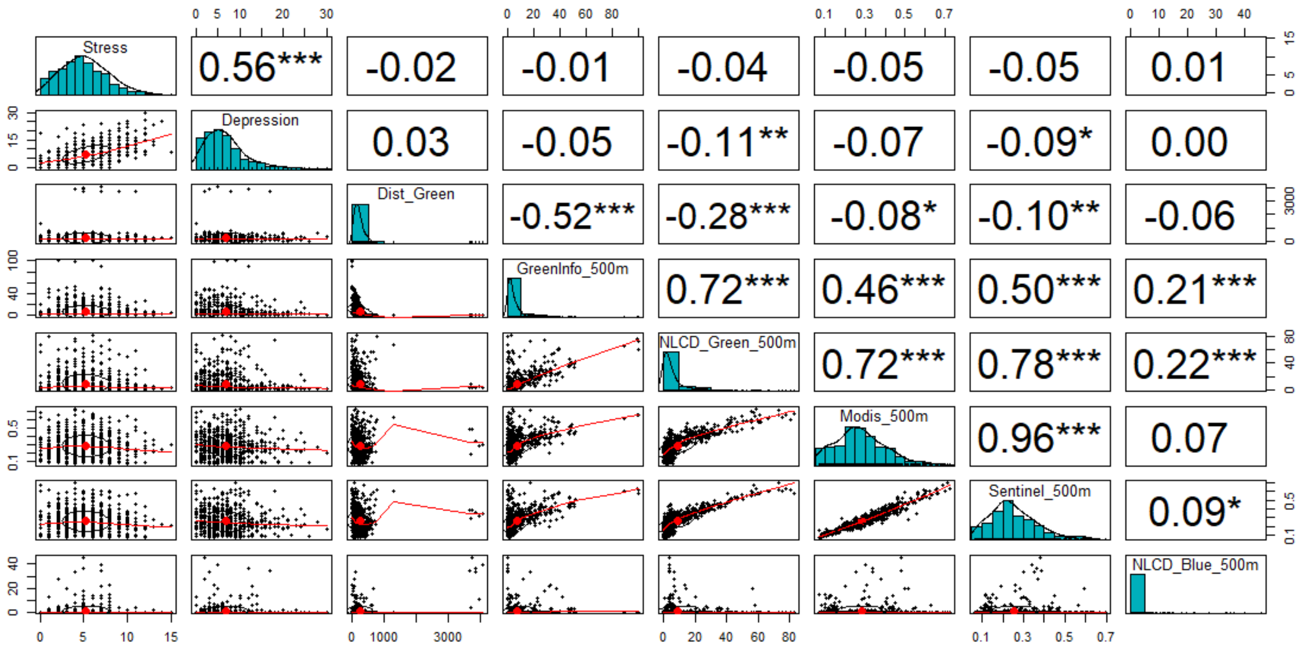


Figure 12. Spearman's correlation scatter plots, histogram, and coefficients between health data and 1000m buffer size green and blue space variables.

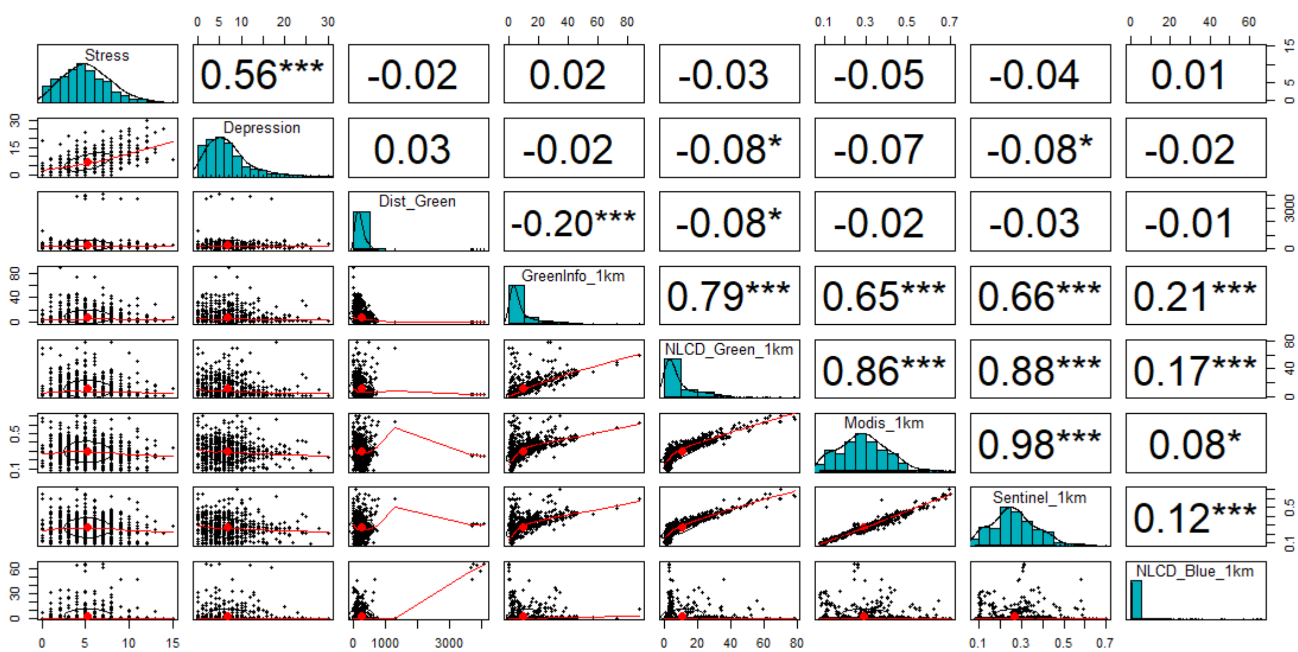
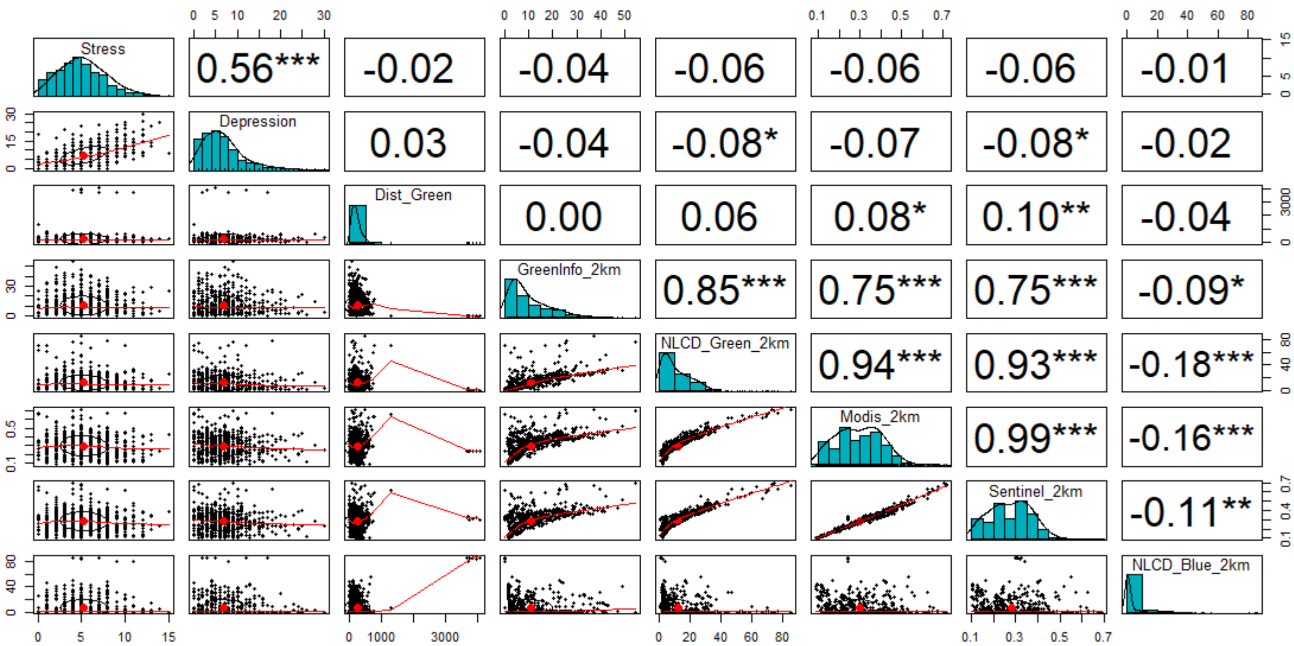


Figure 13. Spearman's correlation scatter plots, histogram, and coefficients between health data and 2000m buffer size green and blue space variables.



3.2 Average levels of green and bluespaces stratified by perceived stress and depression

There were 638 patients assigned with low scores of perceived stress (e.g. scores smaller than 9) and 83 patients with high scores of perceived stress (e.g. scores greater or equal to 9). For the depression dataset, there were 637 patients assigned with low scores of depression (e.g. scores smaller than 16) and 47 patients with high scores of depression (e.g. scores greater or equal to 16). Table 3 illustrates the median, 25th, and 75th percentile values of environmental variables in each sub-group. Figures 14 and 15 are the representation of each of those groups in boxplots.

Comparing the values of environmental variables for low and high sub-groups, I notice that the majority of greenspace values were higher among women with low stress or depression compared to those with high stress or depression (i.e. high greenspace/bluespace values were higher in the low stress/depression groups). This is much more evident on the boxplots where the distributions for the low stress/depression group are clearly higher than the distribution for the high stress/depression group (Figures 14 and 15).

Table 3 shows the low stress group lived farther from parks on average than the high stress group (median of 191m vs. 180m), but that the low depression group lived slightly closer to parks on average than the high depression group (median of 190m vs. 193m). Moreover, the one-sided Wilcoxon Test was used to test the null hypothesis of no difference between the low and high clusters, which was not significant for both datasets: perceived stress and depression.

Taking into account the greenspace buffers in Table 3 (GreenInfo Network, NLCD green, MODIS, and Sentinel), the low stress group had higher average values of greenspace than for high-stress group (except for GreenInfo 300m), with all differences being statistically significant at $P<0.10$ (except for the GreenInfo 300m, 500m, and 2km buffers). A similar pattern was seen with depression, with higher values of greenspace for the low depression sub-group than for the high depression sub-group. Moreover, for every buffer size of NLCD green, MODIS and Sentinel I can observe in the boxplots (figures 14 and 15) that the lower stress/depression sub-groups display in a higher position compared to the high stress/depression sub-groups, indicating more greenspace for the low sub-groups.

The GreenInfo Network buffers (percent within the buffer zone of open access greenspace) were not as consistent as the other measurements. For the perceived stress dataset, the smaller buffer (300m) held a greater percentage of greenspace for the high-stress than the low-stress sub-group. Notice that only the GreenInfo 1km buffer had a statistically significant difference in the mean between the low stress and high-stress sub-groups at $P<0.10$. The depression dataset presented a very similar result, where all buffers but GreenInfo 300m had a greater percentage of greenspace for the low depression than the high depression sub-group. And again, only the means of GreenInfo 1km were significantly different from low and high depression sub-groups.

The vast number of zeros for the NLCD blue variables can be noticed in Table 3, Figures 14 and 15, where 75% of the patients within 300m and 500m buffers had no water bodies. Notice that for the 300m, 500m, and 1km buffers I report the 95th percentile, 75th percentile, and the maximum values of the percentage of water within the buffer area.. Table 3 indicates a more bluespace within 1km among participants with low depression, and more blue spae within all buffer distances considered among the low perceived stress group.

The one-sided Wilcoxon Test was used to test the null hypothesis of no difference between the low and high sub-groups. GreenInfo 1km and every buffer size of NLDC green, MODIS, and Sentinel had a significant p-value. I consider a confidence interval of 90% and an alpha value of 10%. The perceived stress outcome appeared as significantly correlated with 18 environmental variables with p-values lower than 0.1, they were: GreenInfo 1km, NLCD Green 300m, NLCD Green 500m, NLCD Green 1km, NLCD Green 2km, MODIS 300m, MODIS 500m, MODIS 1km, MODIS 2km, Sentinel 300m, Sentinel 500m, Sentinel 1km, Sentinel 2km, NLCD blue 300m, NLCD blue 500m, NLCD blue 1km, and NLCD blue 2km (Table 3).

The depression outcome appeared as significantly correlated with 14 environmental variables, where p-values were lower than 0.1 rejecting the null hypothesis of no difference between the low and high: GreenInfo 1km, NLCD Green 300m, NLCD Green 500m, NLCD Green 1km, NLCD Green 2km, MODIS 300m, MODIS 500m, MODIS 1km, MODIS 2km, Sentinel 300m, Sentinel 500m, Sentinel 1km, Sentinel 2km, and NLCD blue 1km (Table 3).

Table 3. Low and high values of perceived stress and depression. Values of median, 25th, and 75th percentile of green and bluespace variables within each of those groups.

Median(25th, 75th) Environment Variables	PERCEIVED STRESS			DEPRESSION		
	Low Stress (< 9)* 638 patients	High Stress (>= 9)* 83 patients	P-values ˣ	Low Depression (< 16)* 637 patients	High Depression (>= 16)* 47 patients	P-values ˣ
Distance to GreenInfo (m)	191.64 (101.38, 289.95)	180.46 (118.37, 263.63)	0.59	190.80(103.04, 286.42)	193.10(110.83, 284.39)	0.47
GreenInfo 300m %	1.94 (0.08, 5.77)	2.10 (0.28, 4.44)	0.42	2.03(0.11, 5.47)	1.32(0.12, 4.51)	0.37
GreenInfo 500m %	2.91 (1.02, 7.80)	2.46 (1.20, 7.11)	0.21	2.87(1.04, 7.74)	2.64(0.99, 5.64)	0.19
GreenInfo 1km %	4.99 (2.37, 12.21)	3.41 (2.34, 9.21)	0.09	4.98(2.37, 12.40)	3.88(2.28, 7.37)	0.08
GreenInfo 2km %	7.43 (3.37, 15.49)	5.74 (2.99, 14.34)	0.11	7.39(3.32, 15.53)	6.85(3.24, 14.27)	0.32
NLCD Green 300m %	2.41 (0.44, 9.38)	1.66 (0.14, 4.11)	0.02	2.28(0.44, 9.23)	1.17(0.06, 3.61)	0.01
NLCD Green 500m %	3.95 (1.14, 10.92)	2.88 (1.02, 6.70)	0.05	3.89(1.15, 11.54)	1.80(0.78, 4.60)	0.01
NLCD Green 1km %	5.37 (2.59, 16.67)	3.54 (2.23, 11.11)	0.01	5.37(2.48, 16.75)	3.59(2.32, 8.43)	0.01
NLCD Green 2km %	9.37 (3.38, 18.55)	5.88 (2.68, 16.44)	0.07	9.57(3.34, 18.61)	5.87(2.75, 15.09)	0.07
Modis 300m (NDVI)	0.26 (0.20, 0.37)	0.24 (0.17, 0.33)	0.02	0.26(0.19, 0.37)	0.24(0.20, 0.32)	0.10
Modis 500m (NDVI)	0.27 (0.20, 0.37)	0.24 (0.18, 0.33)	0.01	0.27(0.20, 0.37)	0.25(0.20, 0.32)	0.06
Modis 1km (NDVI)	0.29 (0.21, 0.37)	0.25 (0.20, 0.34)	0.01	0.29(0.20, 0.38)	0.27(0.21, 0.31)	0.04
Modis 2km (NDVI)	0.30 (0.21, 0.38)	0.25 (0.20, 0.37)	0.04	0.30(0.21, 0.39)	0.25(0.21, 0.36)	0.07
Sentinel 300m (NDVI)	0.23 (0.17, 0.31)	0.21 (0.16, 0.28)	0.01	0.23(0.17, 0.32)	0.21(0.16, 0.26)	0.03
Sentinel 500m (NDVI)	0.24 (0.18, 0.33)	0.22 (0.17, 0.28)	0.01	0.24(0.18, 0.33)	0.21(0.17, 0.28)	0.02
Sentinel 1km (NDVI)	0.26 (0.21, 0.34)	0.24 (0.19, 0.30)	0.02	0.26(0.20, 0.35)	0.25(0.20, 0.28)	0.03
Sentinel 2km (NDVI)	0.29 (0.21, 0.35)	0.24 (0.19, 0.34)	0.05	0.29(0.21, 0.36)	0.24(0.21, 0.33)	0.07
NLCD Blue 300m %	1.71 (0.00, 30.11)ᵻ	0.00 (0.00, 1.37)ᵻ	0.10	1.39(0.00, 30.11)ᵻ	0.00(0.00, 11.28)ᵻ	0.16
NLCD Blue 500m %	4.99 (0.00, 45.81)ᵻ	0.11 (0.00, 3.72)ᵻ	0.04	4.06(0.00, 45.81)ᵻ	1.05(0.00, 34.84)ᵻ	0.16
NLCD Blue 1km %	17.67 (0.40, 66.37)ᵻ	11.51 (0.09, 46.63)ᵻ	0.10	16.93(0.40, 66.37)ᵻ	7.59(0.03, 61.90)ᵻ	0.08
NLCD Blue 2km %	0.15 (0.01, 7.30)	0.08 (0.01, 3.20)	0.10	0.15(0.01, 7.00)	0.09(0.01, 4.00)	0.31

ˣ All variables were not normally distributed, therefore a one-sided Wilcoxon Test was used to test the null hypothesis of no difference between the low and high groups. *** Values determinate according to Eick et al. (2020).** ᵻ95th percentile (75th percentile, maximum).
Blue shading are the significant p-values (P<0.10).

Figure 14. Box plot of low and high perceived stress for each environmental variable.

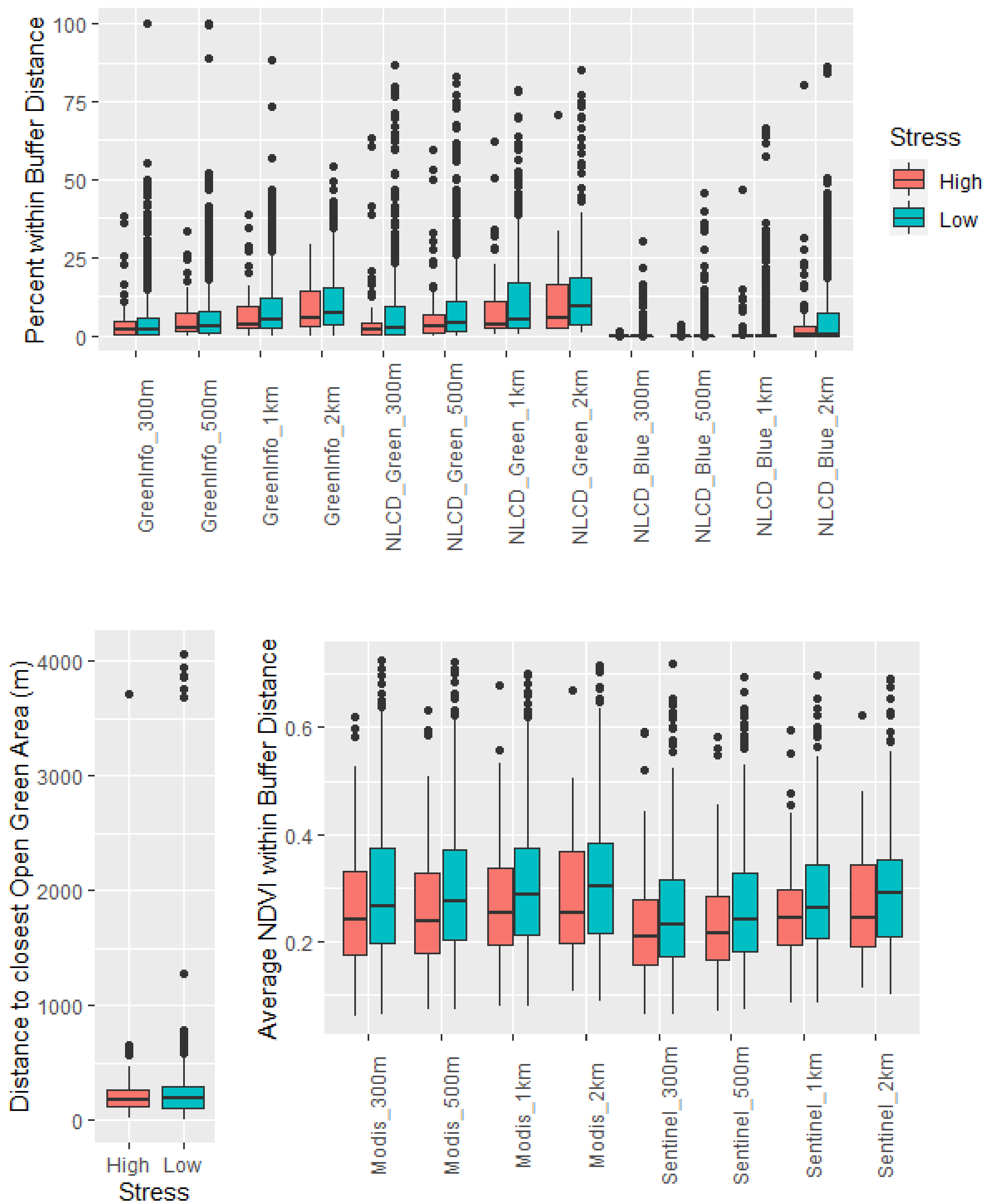
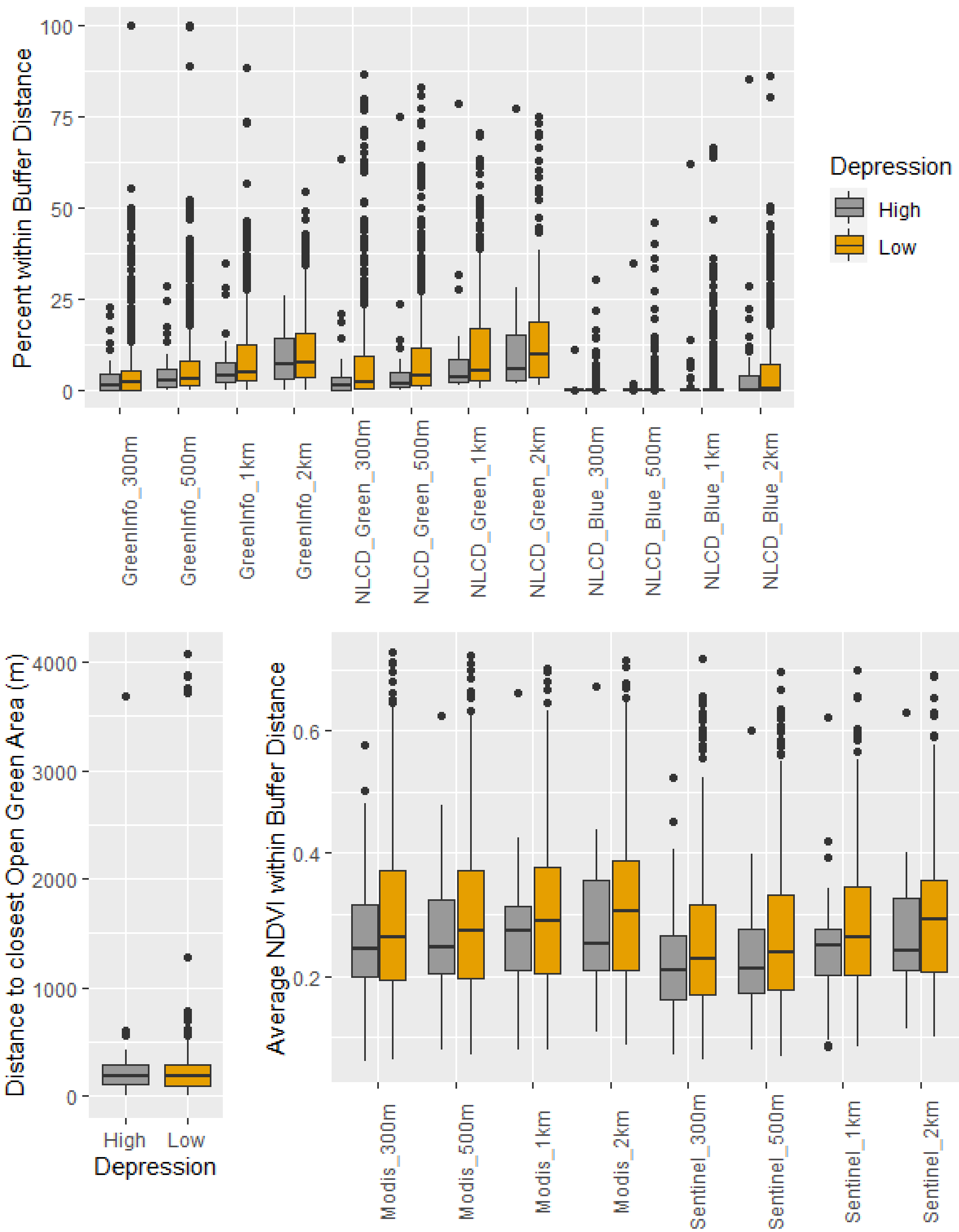


Figure 15. Box plot of low and high depression for each environmental variable.



3.3 Perceived stress and depression scores across greenspace quartiles

The first quartile of distance to the GreenInfo green spaces represented the 25% of patients who lived closest to green spaces and the fourth quartile the 25% of patients who lived furthest away from green spaces. Table 4 shows the mean perceived stress and depression scores across these quartiles. No consistent pattern was seen between perceived stress and distance to parks. For the depression dataset, patients living the closest to parks (Q1) had lower average depression scores (mean = 6.39) than patients living the furthest from the green spaces (Q4 mean = 7.46); however, the values did not decline gradually with further distances to GreenInfo: the second quartile had the highest average depression scores (mean = 7.50).

Taking into account the measurements of greenspace using buffers (GreenInfo network, NLCD green, MODIS, and Sentinel), the first quartile represents a cluster of the lowest concentration of greenspace and the fourth quartile a cluster of the highest values of it. Where GreenInfo and NLCD green buffers were based on the percentage of greenspace within the patient's buffer, and MODIS and Sentinel buffers were the average of NDVI within the patient's buffer area. I calculated the average of perceived stress and depression for each of these quartiles. Overall, the mean values of perceived stress and depression gradually decrease across quartiles of greenspace. Except for GreenInfo 1km, perceived stress was higher for the first quartile than the fourth quartile of greenspace values. In other words, the patients with lower perceived stress were exposed to more greenspace than the ones with higher perceived stress. No variable showed a consistent decrease across all four quartiles; however, the majority of them show a trend of higher values on Q1 and lower values among Q2 or Q3 and Q4. All greenspace variables using buffer zones for the depression dataset showed higher first quartile values than the fourth quartile of the depression mean scores. However, only Sentinel 300m had declining values from all four quartiles;

in other words, the average depression decreased with the increase in greenspace. Despite that, the majority of the variables show a trend of higher values on Q1 and smaller among Q2 or Q3 and Q4.

The one-sided Wilcoxon Test was used to test the null hypothesis of no difference between the means of perceived stress/depression among the first quartile and fourth quartiles. I considered a confidence interval of 90% and an alpha value of 10%. Perceived stress presented one environmental variable with p-values lower than 0.10, which was NLCD green 2km. In another context, the depression dataset had all variables except for GreenInfo 1km with a significant p-value. These variables rejected the null hypothesis of no difference between the mean scores of the first quartile (Q1) and fourth quartiles (Q4).

Table 1. Perceived stress and depression scores across quartiles of greenspace

Greenspace variables clustered as categorical quartiles	Mean Perceived Stress Score across quartiles of greenspace (scores from 0 – 15)					Mean Depression Score across quartiles of greenspace (scores from 0 – 30)				
	Q1	Q2	Q3	Q4	P-values [×]	Q1	Q2	Q3	Q4	P-values [×]
Distance to GreenInfo (m)	5.22	5.43	5.16	5.21	0.52	6.39	7.50	6.50	7.46	0.01
GreenInfo 300m %	5.25	5.22	5.62	4.93	0.12	7.45	6.84	7.39	6.18	0.01
GreenInfo 500m %	5.27	5.24	5.34	5.17	0.37	7.22	6.98	7.22	6.43	0.04
GreenInfo 1km %	5.09	5.28	5.48	5.16	0.71	6.82	7.37	7.36	6.30	0.31
GreenInfo 2km %	5.40	5.26	5.18	5.18	0.17	7.13	7.13	7.21	6.39	0.08
NLCD Green 300m %	5.25	5.50	5.38	4.88	0.12	7.08	7.70	7.34	5.74	0.02
NLCD Green 500m %	5.12	5.72	5.31	4.87	0.24	7.40	7.78	6.87	5.80	0.00
NLCD Green 1km %	5.32	5.37	5.40	4.92	0.16	7.36	7.54	7.16	5.79	0.02
NLCD Green 2km %	5.61	5.32	4.92	5.17	0.06	7.40	7.52	6.58	6.35	0.05
Modis 300m (NDVI)	5.22	5.55	5.21	5.04	0.33	6.91	7.48	7.50	5.96	0.03
Modis 500m (NDVI)	5.31	5.39	5.36	4.95	0.16	6.89	7.79	7.16	6.01	0.06
Modis 1km (NDVI)	5.23	5.42	5.47	4.91	0.14	6.82	7.77	7.29	5.96	0.05
Modis 2km (NDVI)	5.42	5.47	4.94	5.18	0.19	7.03	7.84	6.77	6.22	0.07
Sentinel 300m (NDVI)	5.50	5.33	5.09	5.09	0.12	7.36	7.34	6.84	6.32	0.03
Sentinel 500m (NDVI)	5.34	5.36	5.32	5.00	0.15	7.27	7.53	7.16	5.89	0.01
Sentinel 1km (NDVI)	5.28	5.20	5.56	4.98	0.18	6.95	7.61	7.42	5.88	0.04
Sentinel 2km (NDVI)	5.36	5.49	5.07	5.09	0.19	6.91	8.16	6.71	6.07	0.07

[×] All variables were not normally distributed, therefore a one-sided Wilcoxon Test was used to test the null hypothesis of no difference between the Q1 and Q4 groups. Blue shading are the significant p-values (P<0.10).

The house, arrows and trees are an illustration of the distance to GreenInfo – lower quantiles are closer to greenspace. And the agglomerate of trees are to illustrate the amount of greenspace in each quartile – lower quartiles contain lower amount of greenspace.

4. Discussion

This study explored different methods to measure green and bluespace in order to assess their relationship with perceived stress and depression of pregnant women in the San Francisco Bay Area. I had three main questions: (1) how are self-reported stress and depression levels of pregnant women in the San Francisco Bay Area related to their proximity to green and blue space, (2) which measures of green space are most strongly correlated with stress and depression, and (3) if the spatial resolution of greenspace measures matter.

To answer these questions, I analyzed the data in three ways. First, I calculated the correlation between the various variables. Second, I categorized the data into two sub-groups (low and high perceived stress and depression) to evaluate the amount of blue and greenspaces across them. And third, I calculated the mean perceived stress and depression scores across each greenspace categorical quartile.

4.1 Spearman's correlation discussion analysis

As expected in the hypothesis, Spearman's correlation presented a negative correlation with greenspace and either of the outcomes; however, these values were very small. For measurements of 2km buffers of NLCD, MODIS, and Sentinel-2, the perceived stress dataset proves the hypothesis of low stress being associated with more greenspace. NLCD 500m buffer had the strongest correlation with depression outcomes, proving that more greenspace is related to lower pregnant women's depression levels. From other studies, green and blue spaces may be associated with perceived stress and depression (Alcock et al., 2014; Beyer et al., 2014; Brooks et al., 2017; Gascon et al., 2015; Triguero-Mas et al., 2015; Wheeler et al., 2012; White et al., 2013) and this analysis goes along with that, since there is an association with greenspace. I expected a small association since there are many other factors that influence perceived stress and depression outcomes. Moreover, this analysis considered only the monotonic relationship between the variables (Spearman's correlation). There is the possibility of relationships between them being non-linear and not captured by Spearman’s correlation, which raises questions for further research.

Due to the fact that the distance to GreenInfo and GreenInfo Network 300m and 500m buffers are based on the same dataset, I expected a high correlation between them. That did not happen for larger buffers because the majority of patients lived within a 500m distance to a green area. Notice that the percent of areas of open access green space for GreenInfo buffers were moderate to highly correlated with NLCD. Since I included developed open space into the NLCD greenspace, I believe that the city parks and recreational areas were well represented in both datasets (NLCD and GreenInfo Network), which justifies their correlation.

Considering only the correlation among the environmental samples, there were a few variables that can be considered the most effective in order to represent greenspace in this research context. As expected, Sentinel is highly correlated to MODIS, since both data calculate NDVI for the same month but at different resolutions. Notice that the correlation gets stronger with larger buffer sizes. This can be explained by the number of pixels inside the buffer. For the 300m buffer, MODIS had intersected about 4, 5, or 6 pixels to calculate the average NDVI and Sentinel about 2.800 pixels. On the other hand, for the 2km buffer size, MODIS intersected an average of 230 pixels and Sentinel about 120.000 pixels. For the small buffers, a single missing pixel for the MODIS buffer would present a significant change in the statistics for that patient, which was not true for larger buffer sizes; therefore, the correlation between the two datasets increased with the buffer size.

Moreover, MODIS and Sentinel were very much correlated with NLCD for larger buffer distances and moderate to smaller buffers. That was expected since NLCD data is driven from Landsat imageries and NDVI is one of the methods applied to classify the land cover map. Therefore, larger buffers include more areas that might be more similar to the NDVI.

For this analysis on Spearman's correlation, bluespace had no conclusive correlation with either of the health variables: depression and perceived stress. Although it is common in the greater Bay Area for people who live far from the water to have views of it, proximity to bluespace was only considered for patients living close to water (up to 2km), excluding the possibility of patients who live in homes that had ocean or lake views, which should be explored in further research. Once I take all patients into account, many more have access to greenspace than to bluespace. Therefore, the number of patients with proximity to bluespaces in contrast to the ones who lacked such proximity was minimal, and not noticed in the correlation.

4.2 Average levels of green and bluespaces stratified by perceived stress and depression

I hypothesized for the distance to GreenInfo Network greenspaces that patients closer to such areas would report less perceived stress and depression, and patients further from these areas would report higher perceived stress and depression. However, the results from the analysis using low and high sub-groups did not fully support this hypothesis. In some cases, the sub-group with low perceived stress and depression lived further from greenspaces. On top of that, the p-values were not significant and I could not assume that the two groups were different. Notice that the GreenInfo Network dataset did not represent private lands. Some patients may have lived further from parks but have a nearby green area that they could see or access; therefore, the distance to public open space is not an accurate way to represent greenspace in this health context. The results follow the findings from Tamosiunas et al., (2014), who did not find an association between cardiovascular health and distance to greenspace. On the other hand, Zijlema et al., (2017) established an association between distance to the natural outdoor environment and cognitive function. Considering these factors, I suggest further research to analyze the relationship between low/high groups of mental health and distance to public greenspaces.

Regarding the measurement of greenspace using raster datasets (NLCD green, MODIS, and Sentinel) I observed the consistency across all buffer sizes of higher values of greenspace for the low-stress sub-group than the high-stress sub-group. Different resolutions of the raster images did not affect the fact that the low sub-group had more nearby greenspace than the high-stress sub-group; our hypothesis was incorrect.

Once again considering the measurements of greenspace using raster datasets, depression had a similar result to perceived stress. Overall, we found higher values of greenspace for the low depression sub-group than for the high depression sub-group, except for the MODIS 300m and 1km. These two exceptions had a small number of patients (25th percentile) who presented less greenspace for the low-depression sub-group than the high-depression sub-group. This could suggest a small sensitivity to image resolution regarding depression; however, once analyzing the Wilcoxon Test, the difference between the mean of the low and high depression sub-groups for MODIS were significant for all buffer sizes, including the 300m and 1km.

A valuable finding of this analysis that arguments the findings of other literature investigating greenspace/health associations is that high perceived stress and depression both have a reliable significant association with the amount of local greenspace at all four distances evaluated in this study

Considering these facts and the context of this research, the use of images at different resolutions for the perceived stress and depression datasets did not impact the analysis of greenspace, confirming the findings of Reid et al. (2018).

The results for the GreenInfo Network buffer for perceived stress and depression were not consistent with the other buffers' measurements. The only buffer zone with a significant result was Green Info 1km. Notice that GreenInfo is a vector-based dataset, which can be considered more accurate than the other datasets in representing parks and recreational areas. However, this dataset excludes other greenspaces that appear on NLCD and NDVI measures, such as private greenspaces. Therefore, it appears the use of a dataset with information beyond parks and recreational areas (such as private land) is essential to assess the influence of greenspace on a patient's overall health (perceived stress and depression).

Concerning what was mentioned about private lands, NLCD could exclude areas that appear to be smaller than the pixel size (30m resolution) such as yards, small gardens, or street trees; however, it included private lands (larger than the 30m pixel size) such as golf courses. Still, it was not necessary to capture small gardens, trees, and private yards to conclude that the low sub-group for perceived stress and depression had more nearby greenspace than the high sub-group for perceived stress and depression.

Regarding the NLCD bluespace statistics for low and high sub-groups, both datasets had a vast amount of patients with no water close to their house. I found sensitivity to bluespace in the analysis for low and high sub-groups of perceived stress and depression, especially perceived stress. All NLCD blue buffers for perceived stress had a significant difference between the low and high sub-groups, and the statistics showed that the low perceived stress sub-group had a higher amount of water nearby than the high perceived stress sub-group. On the other hand, the depression dataset had one variable (NLCD blue 1km) that presented these results. It is important to consider that in the greater Bay Area, mostly the neighborhoods in the coastal areas are populated by wealthy families. Therefore, the fact that the stress is more sensitive to being closer to water is consistent with the fact that wealthy people tend to get less stress than the low-income population; however, wealthy families may suffer from depression diseases despite their economic situation. For this reason, I suggest more research associating the socio-economic factors into the context of health analysis and green/bluespace.

4.3 Perceived stress and depression scores across greenspace quartiles

Regarding the categorical analysis, the increased distance to GreenInfo greenspaces did not result in lower scores of perceived stress. On the other hand, patients who lived further away from green areas (quartile four) did show higher scores in depression; however, the depression dataset was not consistent throughout all quartiles. Perceived stress had no significant p-values regard the difference between the mean of Q1 and Q4 for the distance to GreenInfo. In contrast to that, the depression dataset had a significant p-value, which confirms that patients living closer to greenspaces resulted in lower levels of depression.

For the measurements of greenspace using buffers (GreenInfo Network, NLCD green, MODIS, and Sentinel), many variables presented lower scores of perceived stress and depression for Q1 than for Q4. This indicates that patients with a greater concentration of greenspaces in their buffer zone may be less susceptible to stress or depression than patients who had less greenspace within their buffer zone. Moreover, for the depression dataset, the majority of buffer variables presented significant p-values once comparing the mean of Q1 and Q4. This shows that the group of patients living within more greenspace were less depressed than the ones living with a lower amount of greenspace. In contrast to that, perceived stress did not present significant results; only, NLCD green 2km had a significant p-value comparing the mean of Q1 and Q4.

Therefore, categorical data in the context of this research was very consistent in showing that depression was influenced by greenspace when comparing the first and last quartiles. In addition to that, the use of categorical quartiles using buffer zones appears to be more effective for depression than perceived stress.

5. Conclusions

This research analyzed the relationship between the environment (green and bluespace) and the self-reported mental health of pregnant women in the greater Bay Area, California. I applied different methods to calculate the green and bluespace in order to identify the most effective representation of it in this context. Moreover, I analyzed how the level of perceived stress and depression related to these environmental variables.

By using three different analyses, this research has found multiple discoveries for the effect of environment variables on mental health. Based on Spearman's correlation analysis I found a small correlation between mental health and greenspace, with few greenspace variables reaching statistical significance. Considering the analysis of low and high sub-groups of perceived stress and depression, this thesis has shown multiple variables represented the greenspace accurately. For the perceived stress and depression dataset, each one of the buffer sizes was adequately represented by the NLCD, NDVI Modis, and NDVI Sentinel. For the GreenInfo Network, the distance to GreenInfo greenspace was not significant for both datasets and only GreenInfo 1km buffer zone was significant within the difference between the low and high sub-groups. Further, in the same analysis of low and high sub-groups, perceived stress was more sensitive to the percentage of bluespace within the buffer zones than depression. All NLCD blue buffers were significant to prove the difference between low and high sub-groups for perceived stress and only NLCD blue 1km was significant for depression. The outcomes of perceived stress being more sensitive to bluespace than depression can be explained by the fact that is more expensive to live in coastal areas in the greater Bay Area, and wealthy people tend to get less stressed than the low-income population. However, wealthy families may be affected by depression diseases despite their economic situation. This raises an important socio-economic question regardless of the effect of greenspace and health; therefore, further research should investigate that.

We could observe in the analysis of low and high sub-groups of mental health data that the use of different spatial resolutions had a very small influence on the health parameters. This is suggesting that moving to the finer resolution of Sentinel-2 for greenspace may not be important. For perceived stress, image resolution (among MODIS, Sentinel-2, and Landsat-derived NLCD) did not affect the results. For depression, there was a small difference between the low and high sub-groups regard to the MODIS 300m and 1km buffer, which did not occur with the other buffer sizes. As a result, NLCD, Sentinel, and MODIS were great sources to represent the greenspace in this context. They capture greenness that sometimes the subject might not have access to, nevertheless, they still might see it, which can increase their overall health.

The analysis of categorical green quartile data for the perceived stress dataset did not present significant results once comparing the mean perceived stress of the first and fourth quartiles of greenspaces. In contrast, the use of categorical green quartiles for the depression data source was extremely significant for the majority of greenspace variables, including GreenInfo Network, NLCD, MODIS, and Sentinel.

The use of categorical quartiles for the depression dataset was the only consistent analysis that proves distance to GreenInfo and the GreenInfo buffer zones to be significant when comparing the mean depression for the first and fourth quartiles. Once comparing these results with the other statistical analyses, the use of vector data (GreenInfo Network) was not efficient for assessing distance to the closest greenspace and percentage of greenspace within the buffer zone in this health context. Consequently, further research is needed to determine the relationship between green data from vector databases and mental health.

To better understand the implications of these results, future studies could address the question of whether subjects do not actually need to access green or bluespaces to increase their overall health. Within the Bay Area, a lot of people who lack access to the ocean or parks within a 2km radius nonetheless might have a panoramic view of it. Additionally, new studies could address questions to compare views are to direct experience of natural environments.

This research analyses greenspace access by Euclidian distance to the closest source and within a buffer area. I suggest more research using street network analysis instead of Euclidian distances since I did not find a consistent correlation between the health outcomes and distance to greenspace. Moreover, from an economic perspective, there are a number of poor neighborhoods that are located close to the water in the San Francisco Bay Area. Consequently, further research is needed when considering other influences as the quality of the surrounding environment, whether a residence has a natural view, social-economic behavior, food insecurity, financial strain, poor neighborhood quality, among other impacts. Furthermore, the use of buffer distances limited the area where the patient could access their surroundings. To better understand the implication of these results, future studies could address the subjects' own neighborhood descriptions, include their work zone and the areas where they commonly access green/bluespaces.

Based on these conclusions, stress, and depression severely affect people globally and the environment might have some positive effect on people's lives. Consequently, I was able to show correlations between greenspace and mental health during pregnancy, which could have important implications for land use planning and maternal/infant health. Moreover, this is the first time that research has compare different methods to measure the environment and mental health of pregnant women in the greater Bay Area in California.

References

Alcock, I., White, M. P., Wheeler, B. W., Fleming, L. E., & Depledge, M. H. (2014). Longitudinal Effects on Mental Health of Moving to Greener and Less Green Urban Areas. *Environmental Science & Technology*, *48*(2), 1247–1255. https://doi.org/10.1021/es403688w

Beil, K., & Hanes, D. (2013). The Influence of Urban Natural and Built Environments on Physiological and Psychological Measures of Stress—A Pilot Study. *International Journal of Environmental Research and Public Health*, *10*(4), 1250–1267. https://doi.org/10.3390/ijerph10041250

Beyer, K., Kaltenbach, A., Szabo, A., Bogar, S., Nieto, F., & Malecki, K. (2014). Exposure to Neighborhood Green Space and Mental Health: Evidence from the Survey of the Health of Wisconsin. *International Journal of Environmental Research and Public Health*, *11*(3), 3453–3472. https://doi.org/10.3390/ijerph110303453

Brooks, A. M., Ottley, K. M., Arbuthnott, K. D., & Sevigny, P. (2017). Nature-related mood effects: Season and type of nature contact. *Journal of Environmental Psychology*, *54*, 91–102. https://doi.org/10.1016/j.jenvp.2017.10.004

Carrus, G., Scopelliti, M., Lafortezza, R., Colangelo, G., Ferrini, F., Salbitano, F., Agrimi, M., Portoghesi, L., Semenzato, P., & Sanesi, G. (2015). Go greener, feel better? The positive effects of biodiversity on the well-being of individuals visiting urban and peri-urban green areas. *Landscape and Urban Planning*, *134*, 221–228. https://doi.org/10.1016/j.landurbplan.2014.10.022

Casey, J., James, P., Rudolph, K., Wu, C.-D., & Schwartz, B. (2016). Greenness and Birth Outcomes in a Range of Pennsylvania Communities. *International Journal of Environmental Research and Public Health*, *13*(3), 311. https://doi.org/10.3390/ijerph13030311

Coffey, R. (2013, January 18). *The difference between "Land Use" and "Land cover."* MSU Extension – Michigan State University. The Difference between "Land Use" and "Land Cover." https://www.canr.msu.edu/news/the_difference_between_land_use_and_land_cover

CPAD Database Manual. (2019). GreenInfo Network.

Cusack, L. (2017). *Associations between multiple green space measures and birth weight across two US cities*. 8.

Cusack, L., Larkin, A., Carozza, S., & Hystad, P. (2017). Associations between residential greenness and birth outcomes across Texas. *Environmental Research*, *152*, 88–95. https://doi.org/10.1016/j.envres.2016.10.003

Dadvand, P., de Nazelle, A., Figueras, F., Basagaña, X., Su, J., Amoly, E., Jerrett, M., Vrijheid, M., Sunyer, J., & Nieuwenhuijsen, M. J. (2012). Green space, health inequality and pregnancy. *Environment International*, *40*, 110–115. https://doi.org/10.1016/j.envint.2011.07.004

de Vries, S., Verheij, R. A., Groenewegen, P. P., & Spreeuwenberg, P. (2003). Natural Environments—Healthy Environments? An Exploratory Analysis of the Relationship between Greenspace and Health. *Environment and Planning A: Economy and Space*, *35*(10), 1717–1731. https://doi.org/10.1068/a35111

Eick, S. M., Goin, D. E., Izano, M. A., Cushing, L., DeMicco, E., Padula, A. M., Woodruff, T. J., & Morello-Frosch, R. (2020). Relationships between psychosocial stressors among pregnant women in San Francisco: A path analysis. *PLOS ONE*, *15*(6), e0234579. https://doi.org/10.1371/journal.pone.0234579

Erdas Imagine. (2018).

ESRI ArcGIS Pro 2.3.3. (n.d.). ESRI.

Gascon, M., Triguero-Mas, M., Martínez, D., Dadvand, P., Forns, J., Plasència, A., & Nieuwenhuijsen, M. (2015). Mental Health Benefits of Long-Term Exposure to Residential Green and Blue Spaces: A Systematic Review. *International Journal of Environmental Research and Public Health*, *12*(4), 4354–4379. https://doi.org/10.3390/ijerph120404354

Gascon, M., Triguero-Mas, M., Martínez, D., Dadvand, P., Rojas-Rueda, D., Plasència, A., & Nieuwenhuijsen, M. J. (2016). Residential green spaces and mortality: A systematic review. *Environment International*, *86*, 60–67. https://doi.org/10.1016/j.envint.2015.10.013

Glazer, K. B., Eliot, M. N., Danilack, V. A., Carlson, L., Phipps, M. G., Dadvand, P., Savitz, D. A., & Wellenius, G. A. (2018). Residential green space and birth outcomes in a coastal setting. *Environmental Research*, *163*, 97–107. https://doi.org/10.1016/j.envres.2018.01.006

Golden Gate Weather—Monthly Rainfall. (n.d.). Retrieved October 31, 2020, from https://ggweather.com/monthly/sf.htm

Grazuleviciene, R., Danileviciute, A., Dedele, A., Vencloviene, J., Andrusaityte, S., Uždanaviciute, I., & Nieuwenhuijsen, M. J. (2015). Surrounding greenness, proximity to city parks and pregnancy outcomes in Kaunas cohort study. *International Journal of Hygiene and Environmental Health*, *218*(3), 358–365. https://doi.org/10.1016/j.ijheh.2015.02.004

Kondo, M. C., Fluehr, J. M., McKeon, T., & Branas, C. C. (2018). Urban Green Space and Its Impact on Human Health. *Int. J. Environ. Res. Public Health*, 28.

Lachowycz, K., & Jones, A. P. (2014). Does walking explain associations between access to greenspace and lower mortality? *Social Science & Medicine*, *107*, 9–17. https://doi.org/10.1016/j.socscimed.2014.02.023

Laurent, O., Wu, J., Li, L., & Milesi, C. (2013). Green spaces and pregnancy outcomes in Southern California. *Health & Place*, *24*, 190–195. https://doi.org/10.1016/j.healthplace.2013.09.016

Mather, P. M., & Koch, M. (2011). *Computer processing of remotely-sensed images: An introduction* (4th ed). Wiley-Blackwell.

Multi-Resolution Land Characteristics. (n.d.). Retrieved October 31, 2020, from https://www.mrlc.gov/data/type/land-cover

Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(2), 481–486. https://doi.org/10.1109/36.377948

NASA. (n.d.-a). *EARTHDATA SEARCH*. EARTHDATA SEARCH. Retrieved October 31, 2020, from https://search.earthdata.nasa.gov/search

NASA. (n.d.-b). *MODIS Web*. Retrieved October 31, 2020, from https://modis.gsfc.nasa.gov/data/dataprod/mod13.php

Nieuwenhuijsen, M. J., Kruize, H., Gidlow, C., Andrusaityte, S., Antó, J. M., Basagaña, X., Cirach, M., Dadvand, P., Danileviciute, A., Donaire-Gonzalez, D., Garcia, J., Jerrett, M., Jones, M., Julvez, J., van Kempen, E., van Kamp, I., Maas, J., Seto, E., Smith, G., … Grazuleviciene, R. (2014). *P* ositive *h* ealth *e* ffects of the *n* atural *o* utdoor environment in *ty* pical *p* opulations in different regions in *E* urope (*PHENOTYPE*): A study programme protocol. *BMJ Open*, *4*(4), e004951. https://doi.org/10.1136/bmjopen-2014-004951

Oreskovic, N. M., Perrin, J. M., Robinson, A. I., Locascio, J. J., Blossom, J., Chen, M. L., Winickoff, J. P., Field, A. E., Green, C., & Goodman, E. (2015). Adolescents' use of the built environment for physical activity. *BMC Public Health*, *15*(1), 251. https://doi.org/10.1186/s12889-015-1596-6

R version 4.0.2. (n.d.).

Reid, C. E., Kubzansky, L. D., Li, J., Shmool, J. L., & Clougherty, J. E. (2018). It's not easy assessing greenness: A comparison of NDVI datasets and neighborhood types and their associations with self-rated health in New York City. *Health & Place*, *54*, 92–101. https://doi.org/10.1016/j.healthplace.2018.09.005

Rodríguez, D. A., Cho, G.-H., Evenson, K. R., Conway, T. L., Cohen, D., Ghosh-Dastidar, B., Pickrel, J. L., Veblen-Mortenson, S., & Lytle, L. A. (2012). Out and about: Association of the built environment with physical activity behaviors of adolescent females. *Health & Place*, *18*(1), 55–62. https://doi.org/10.1016/j.healthplace.2011.08.020

Rstudio version 1.3.1056. (n.d.).

Sentinel-2—Missions—Sentinel Online. (n.d.). Retrieved October 31, 2020, from https://sentinel.esa.int/web/sentinel/missions/sentinel-2

Tamosiunas, A., Grazuleviciene, R., Luksiene, D., Dedele, A., Reklaitiene, R., Baceviciene, M., Vencloviene, J., Bernotiene, G., Radisauskas, R., Malinauskiene, V., Milinaviciene, E., Bobak, M., Peasey, A., & Nieuwenhuijsen, M. J. (2014). Accessibility and use of urban green spaces, and cardiovascular health: Findings from a Kaunas cohort study. *Environmental Health*, *13*(1), 20. https://doi.org/10.1186/1476-069X-13-20

Triguero-Mas, M., Dadvand, P., Cirach, M., Martínez, D., Medina, A., Mompart, A., Basagaña, X., Gražulevičienė, R., & Nieuwenhuijsen, M. J. (2015). Natural outdoor environments and mental and physical health: Relationships and mechanisms. *Environment International*, *77*, 35–41. https://doi.org/10.1016/j.envint.2015.01.012

USGS. (n.d.). *EarthExplorer*. EarthExplorer. Retrieved October 31, 2020, from https://earthexplorer.usgs.gov/

Villeneuve, P. J., Jerrett, M., G. Su, J., Burnett, R. T., Chen, H., Wheeler, A. J., & Goldberg, M. S. (2012). A cohort study relating urban green space with mortality in Ontario, Canada. *Environmental Research*, *115*, 51–58. https://doi.org/10.1016/j.envres.2012.03.003

Völker, S., & Kistemann, T. (2015). Developing the urban blue: Comparative health responses to blue and green urban open spaces in Germany. *Health & Place*, *35*, 196–205. https://doi.org/10.1016/j.healthplace.2014.10.015

Wheeler, B. W., White, M., Stahl-Timmins, W., & Depledge, M. H. (2012). Does living by the coast improve health and wellbeing? *Health & Place*, *18*(5), 1198–1201. https://doi.org/10.1016/j.healthplace.2012.06.015

White, M. P., Alcock, I., Wheeler, B. W., & Depledge, M. H. (2013). Coastal proximity, health and well-being: Results from a longitudinal panel survey. *Health & Place*, *23*, 97–103. https://doi.org/10.1016/j.healthplace.2013.05.006

Wilker, E. H., Wu, C.-D., McNeely, E., Mostofsky, E., Spengler, J., Wellenius, G. A., & Mittleman, M. A. (2014). Green space and mortality following ischemic stroke. *Environmental Research*, *133*, 42–48. https://doi.org/10.1016/j.envres.2014.05.005

Wolfe, M. K., Groenewegen, P. P., Rijken, M., & de Vries, S. (2014). Green space and changes in self-rated health among people with chronic illness. *European Journal of Public Health*, *24*(4), 640–642. https://doi.org/10.1093/eurpub/cku081

Yin, P. (2019). Comparison of greenness measures in assessing the association between urban residential greenness and birth weight. *Urban Forestry & Urban Greening*, *46*, 126519. https://doi.org/10.1016/j.ufug.2019.126519

Zhan, Y., Liu, J., Lu, Z., Yue, H., Zhang, J., & Jiang, Y. (2020). Influence of residential greenness on adverse pregnancy outcomes: A systematic review and dose-response meta-analysis. *Science of The Total Environment*, *718*, 137420. https://doi.org/10.1016/j.scitotenv.2020.137420

Zijlema, W. L., Triguero-Mas, M., Smith, G., Cirach, M., Martinez, D., Dadvand, P., Gascon, M., Jones, M., Gidlow, C., Hurst, G., Masterson, D., Ellis, N., van den Berg, M., Maas, J., van Kamp, I., van den Hazel, P., Kruize, H., Nieuwenhuijsen, M. J., & Julvez, J. (2017). The relationship between natural outdoor environments and cognitive functioning and its mediators. *Environmental Research*, *155*, 268–275. https://doi.org/10.1016/j.envres.2017.02.017