Consilience: The Journal of Sustainable Development
Iss. 25 (2021-22)

CONSILIENCE

THE JOURNAL OF SUSTAINABLE DEVELOPMENT

Using Spatial Context and Demographic Analysis to Assess Surface Permeability in Philadelphia

Nicholas Zhu
Urban Studies Program
Advised by: Dr. Elaine Simon and Dr. Elizabeth Greenspan
University of Pennsylvania

Abstract

Surface permeability significantly impacts the urban environment. Specifically, impermeable surfaces result in runoff, which in turn causes flooding and pollution. Left unchecked, impermeable surfaces can lead to hazardous conditions for unlucky city residents. These concerns are prominent in Philadelphia, and in response the municipal government has launched an ambitious plan to increase permeability by installing green infrastructure. This paper explores how spatial and demographic research can be combined to provide a holistic assessment of surface permeability across one of America's largest cities. First, 2020 satellite imagery, provided by the United States Geological Survey (USGS), was used to classify permeable and impermeable surfaces over the entire city. Next, demographic data from the 2014-2018 American Community Survey (ACS)– household income, rent, and home value, all by census block group – were individually merged with the surface permeability classification to generate three overlays. For example, surface permeability was correlated with household income . Upon quantitative and qualitative examination of these overlays, it was found that impermeable surfaces are unevenly distributed and inequitably concentrated in Philadelphia's less-affluent communities. Overall, the methodology used in this research demonstrates a multi-disciplinary and reproducible procedure for joint environmental-demographic research. Additionally, the conclusions reached offer location-specific insights that can help inform Philadelphia's future green infrastructure investments and runoff mitigation strategies.

Introduction

Surface permeability is often overlooked. Truth be told, ground observation is usually for avoiding puddles, or worse, the trails of a dog owner without a plastic bag. Even though surface permeability is easily forgotten, it is critical to the environmental health of cities. Impermeable surfaces which include common materials such as concrete, asphalt, gravel, and tar–lead to flooding and stormwater runoff. Runoff is particularly a problem in cities with combined waste-stormwater sewer systems: Philadelphia is such a city. When it rains, these sewers cannot handle the sheer volume of liquid and subsequently release excess untreated waste into local waterways. This process–called combined sewer overflow (CSO)–is severely detrimental to human health and the environment. Flooding is also a major threat to cities, as demonstrated by the devastation wrought by Hurricane Katrina on New Orleans. Permeable surfaces protect against such environmental damage by absorbing stormwater and ensuring that sewers are not overwhelmed. As demonstrated by Hurricane Katrina, flood mitigation infrastructure is not equitably distributed across cities: Poorer neighborhoods are often left more susceptible to stormwater hazards.

This study examines the relationship between surface permeability and socioeconomic factors in order to assess stormwater runoff and environmental justice across Philadelphia. In doing so, this paper asks the reader to consider this question: What would equitable surface permeability in Philadelphia look like? While there is no concrete answer provided in the following pages, this research demonstrates that surface permeability is spatially and demographically unequal across Philadelphia. While it is unrealistic to expect 100 percent permeable surfaces in any urban environment, the dominance of impermeable surfaces in Philadelphia, especially within lower-income communities, is cause for concern.

Literature Review

Many common elements of the environment can be considered green infrastructure (GI), including street trees, lawns, and public parks. Thus, assessing GI requires a detailed survey of land characteristics (Xu et al., 2018). In these surveys, remote sensing observation is more efficient than land-based techniques. Remotely sensed imagery can be collected in many ways, the most common being satellites such as the NASA/USGS Landsat missions. A wide variety of remote sensing techniques have been applied to surface permeability research in urban and suburban settings. Many indices have been developed to identify permeable land characteristics, such as vegetation, in remotely sensed imagery (Labib & Harris, 2018; Okujeni et al., 2018; Padmanaban et al., 2019; Taramelli et al., 2019).

In environmental justice research, spatial context is essential. High resolution geographical analysis enables correlations to be made between environmental characteristics and demographic factors (Weigand et al., 2019). Spatial context is especially important in cities because metropolitan areas have dense and diverse landscapes. In highly populated regions, environmental data are often aggregated over too large of an area. These data then lose their resolution and specificity. Detailed spatial context helps researchers to minimize the ecological fallacy of using group data to draw conclusions about an individual member of the group. Throughout cities, surface permeability and environmental justice are unquestionably linked. Comparatively impermeable neighborhoods experience more flooding and runoff–and suffer from more combined sewer overflows into local waterways. A recent storm that produced one-quarter inch of rain on the ground led to trash and raw sewage flowing down Frankford Creek, which lies in an impermeable semi-industrial region of Philadelphia (Kummer, 2019). Recent studies have reported correlation between surface permeability, exposure to flooding-based contamination, and socioeconomic status (Sansom et al., 2016).

Philadelphia happens to be a national leader in water management. Notably, the city is taking a completely GI-based approach toward addressing their outdated, yet unfortunately typical, combined sewer system (Dolowitz et al., 2018; Fitzgerald & Laufer, 2017). Specifically, in 2011, the municipal government committed to a 25-year $2-billion plan called Green City Clean Waters, which aims to create over 9,500 acres of new permeable surfaces (City of Philadelphia, 2014). The city's proactive decision to focus so heavily on GI implementation was atypical among its peers. In contrast, Washington D.C. proposed an unrealistically expensive gray infrastructure approach that would have separated the sewer system into waste and stormwater components (Bauers, 2012). A New York-based study found that GI can be cheaper than gray infrastructure even without accounting for additional benefits of GI such as heat island mitigation (Montalto et al., 2007).

While Green City Clean Waters is definitely a positive for Philadelphia, stormwater management concerns are still pervasive in the city. For example, the 2019 incident in Frankford Creek left the waterway muddy brown and highly contaminated with hazards such as fecal bacteria (Kummer, 2019). This incident shows that, despite Philadelphia's efforts, a clear need still exists for further GI implementation and surface permeability research. Small storms continue to wreak havoc by introducing hazardous conditions in Philadelphia's communities.

Creating a Supervised Classification

A supervised classification[1] can efficiently identify permeable, impermeable, and water surfaces across Philadelphia. A high-quality satellite image of Philadelphia with minimal cloud cover is required for this process. I downloaded the selected image from the USGS EarthExplorer. The image was taken on September 8, 2020, by the Landsat 8 satellite.[2] Clouds covered 14 percent of the full image but none of Philadelphia. I trimmed the downloaded image and converted it to a raster[3] in which each pixel contained surface reflectance values for ten electromagnetic bands.[4,5] Figure 1 is an image constructed from the three electromagnetic bands in the visible light spectrum (blue, green, and red).[6] This representation mimics what human eyes would see if looking upon Philadelphia from above.

[1] A supervised classification is a type of machine learning task that automatically partitions data based on pre-determined categories.

[2] Landsat 8 was launched in 2013, orbits the Earth every 99 minutes, and collects imagery with 30 meter by 30 meter resolution.

[3] A raster is a matrix of pixels in which each pixel contains specific information.

[4] Specifically, I used the following bands: coastal aerosol, blue, green, red, near infrared, shortwave infrared 1, shortwave infrared 2, cirrus, thermal infrared 1, and thermal infrared 2.

[5] I performed the trimming and raster conversion in ArcGIS.

[6] I performed the image construction in R.

Figure 1. True Color Composite of 2020 Landsat 8 Satellite Imagery

Next, I collected ground-truthed training data and placed them into a shapefile.[7,8] I directly confirmed these individual training points as either permeable, impermeable[9], or water. In total, I confirmed 418 permeable points, 342 impermeable points, and 212 water points for a total of 972 training data points. Figure 2 is a map displaying the locations of all 972 training points.

[7] I created the shapefile in ArcGIS.

[8] While collecting training data for spatial analysis, accurate ground truthing is essential. If, for example, 50 impermeable training points were actually water bodies, the model's ability to classify surfaces would be directly compromised.

[9] In this instance of ground-truthing, impermeable surfaces consisted entirely of vegetation elements. Therefore, permeable elements such as porous pavement were not classified as permeable. This is a limitation of a satellite-imagery based supervised classification.

Figure 2. Training Points Confirmed Through Personal Observation

After collecting the required imagery and training data, I wrote a supervised classification script (see Appendix).[10] Using a decision tree structure, I computed predictions of surface permeability across Philadelphia. Figure 3 shows a generalized example of the decision tree.[11] Figure 4 maps the resulting classification of surface types.

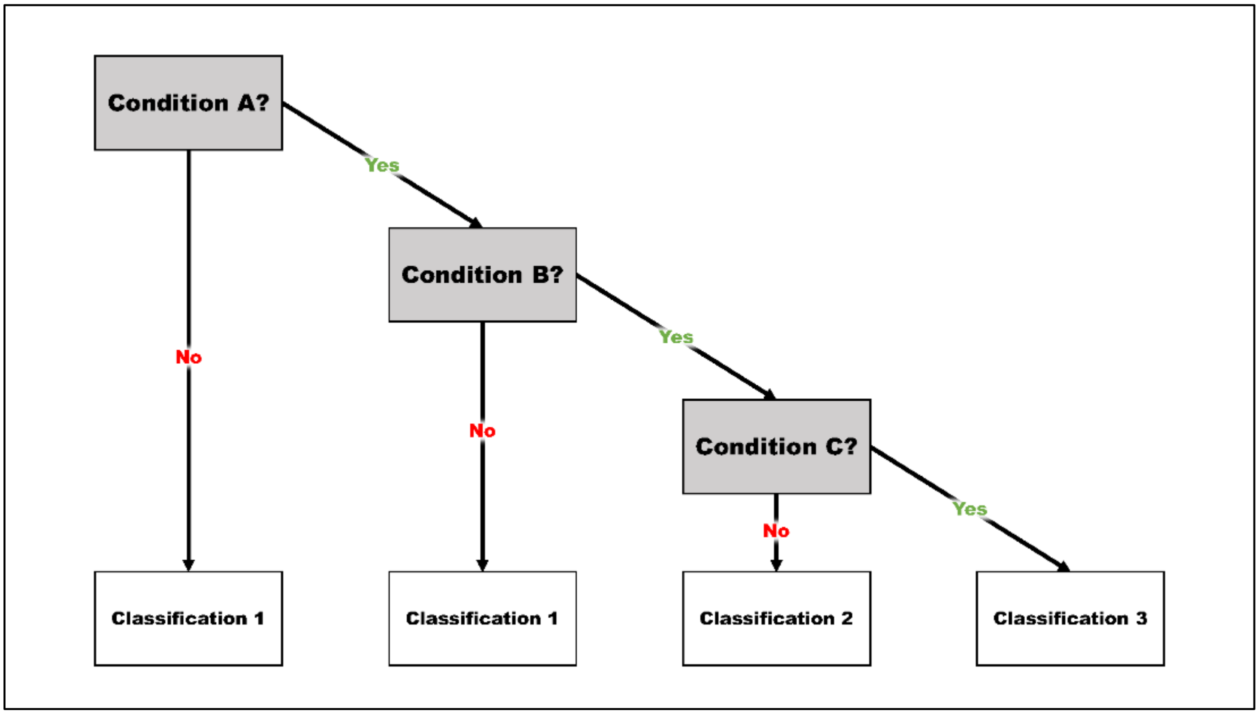


Figure 3. Example of Decision Tree Structure

[10] I wrote the supervised classification script in R.

[11] In the generalized example, each condition represents a reflectance value. The model bases its prediction on the reflectance values of the training points. For example, in Pixel Q, if Band 2's reflectance is greater than X and Band 2's reflectance in permeable-surface training points is almost always greater than X, then Pixel Q is predicted to be a permeable surface.

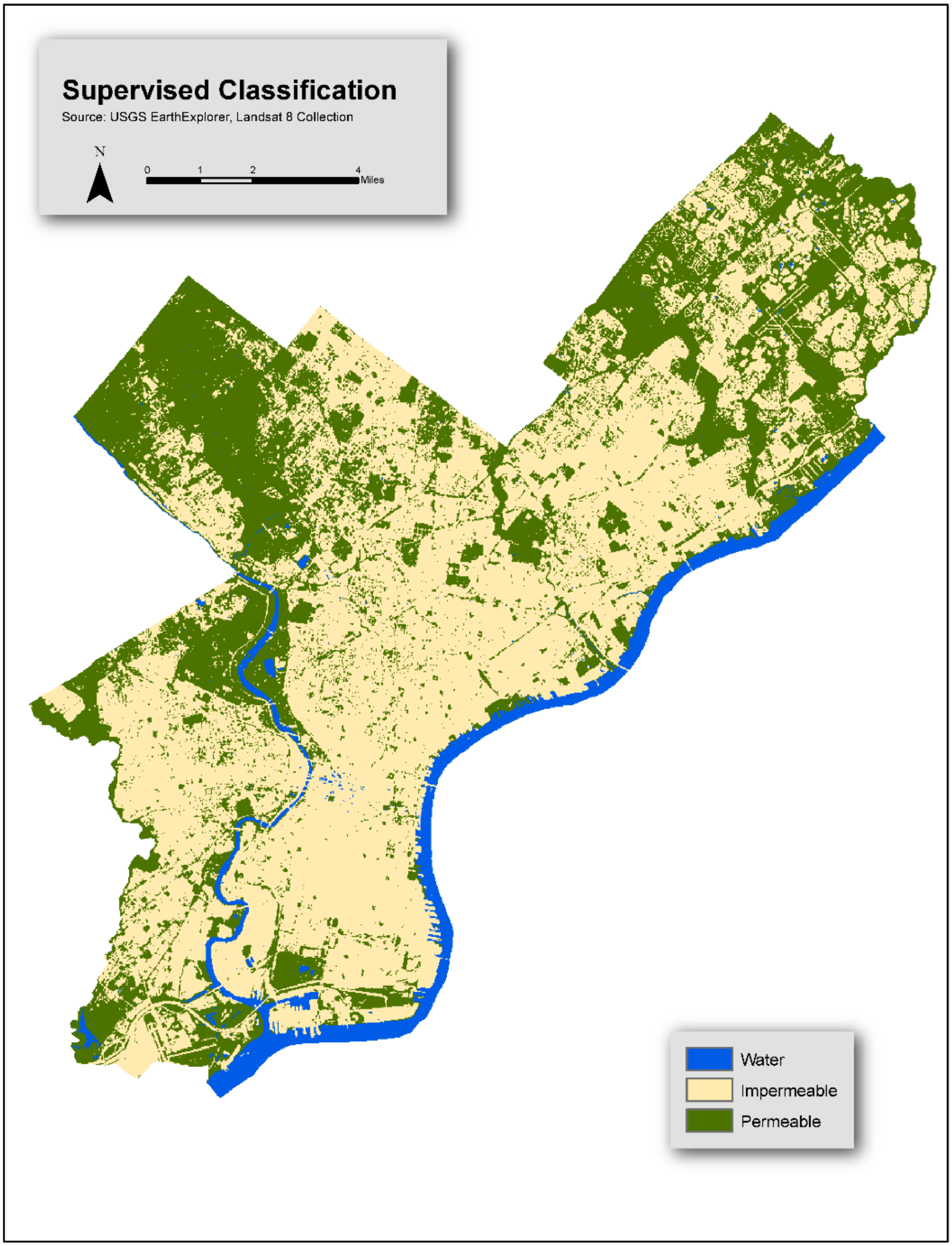


Figure 4. Supervised Classification of Surface Permeability in Philadelphia (2020)

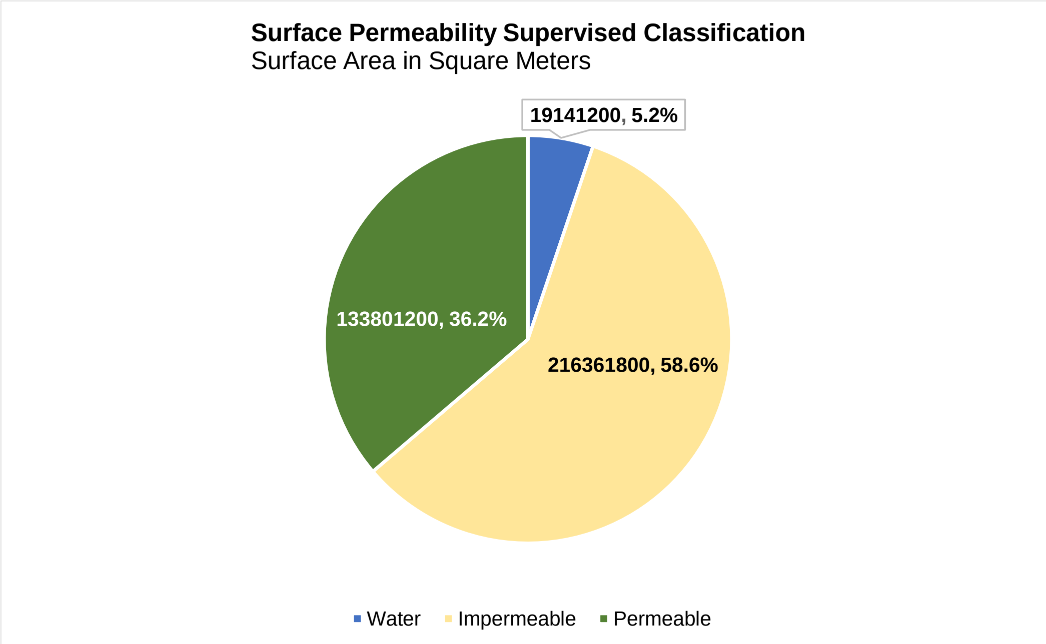


Figure 5. Surface Area Breakdown of 2020 Supervised Classification

As shown in figure 5, impermeable surfaces outnumber permeable surfaces in Philadelphia by nearly 2:1. Impermeable surfaces especially outnumber permeable surfaces in central and south Philadelphia. In the northwest and northeast regions, permeable surfaces are relatively more common. These northern regions represent the sections of Philadelphia that are beginning to blend with suburban landscapes.

To confirm the accuracy of the classification, I compared the predicted permeability with the original training data. The following is an example of this comparison in question form: For a training point that was confirmed in-person as an impermeable surface, was an impermeable surface predicted by the machine learning model? For this criterion, the model was shown to be 99.59 percent accurate: Of the 972 training points, 968 were predicted correctly. One water training point and three permeable training points were incorrectly predicted as impermeable. As another validation method, I calculated the sensitivity[12] and specificity[13] for each category (see table 1).

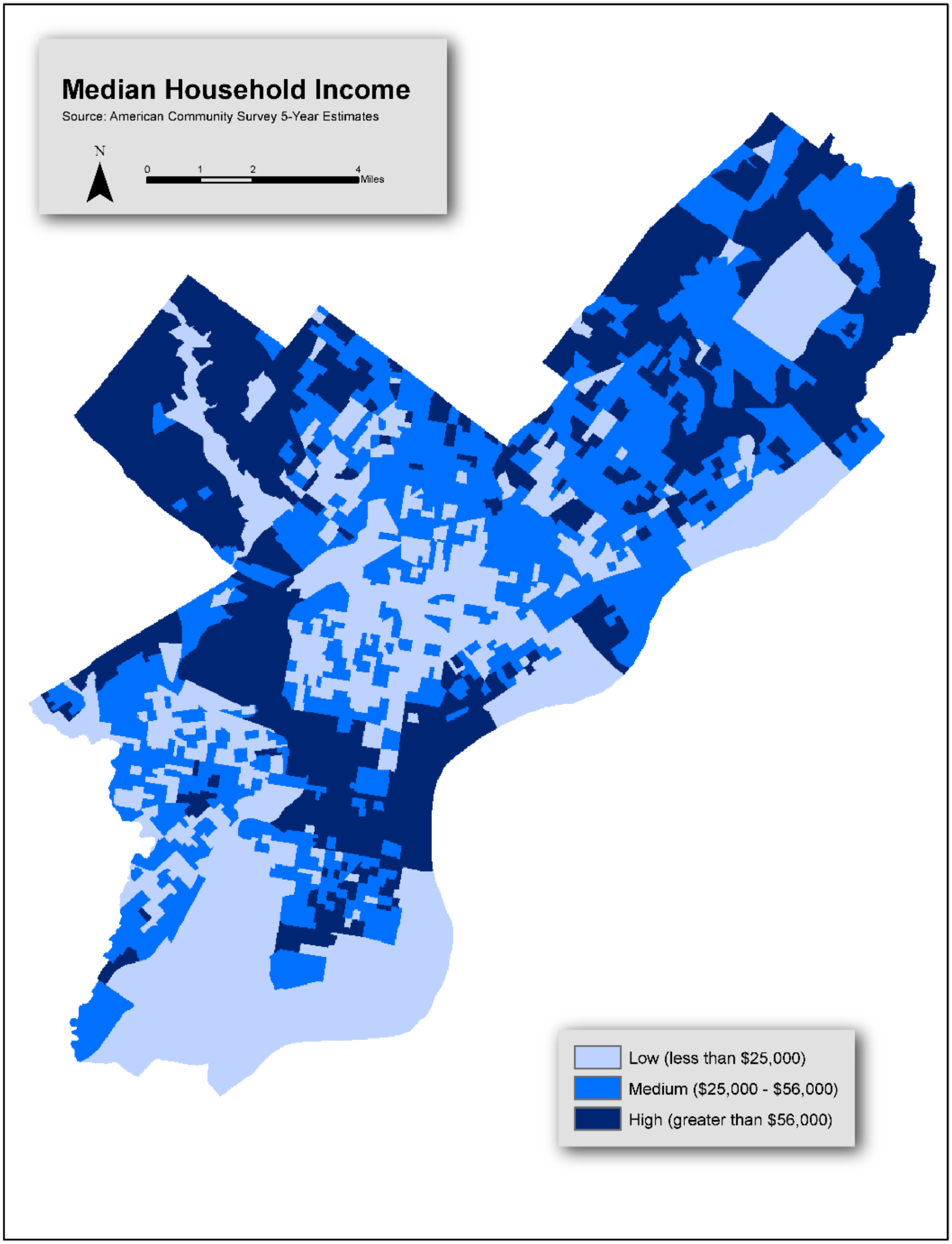
	Permeable	Impermeable	Water
Sensitivity	100.00%	98.84%	100.00%
Specificity	99.46%	100.00%	99.87%

Table 1. Sensitivity and Specificity of the Supervised Classification

Demographic Analysis

This study searches for connections between surface permeability and socioeconomic factors in order to consider issues of environmental justice and neighborhood affordability. Specifically, this study examines a total of three community demographic indicators: household income, rent, and property values.[14] I obtained the census block group for each indicator from the American Community Survey 2014-2018 5 Year Estimates. I then converted this data to raster outputs and separated the block groups into three categories: low, medium, and high (see figures 6, 7, and 8

[12] Sensitivity definition: (number of true positives) / (number of true positives + number of false negatives)
[13] Specificity definition: (number of true negatives) / (number of true negatives + number of false positives)
[14] These indicators, while useful, are not a complete portrait of any community. A limitation of this demographic analysis is that it is impossible to create a complete depiction of an urban community using such indicators.



below). The break points were chosen to create roughly equal amounts of area in each category.[15]Figure 6. Median Household Income from the 2014-2018 American Community Survey

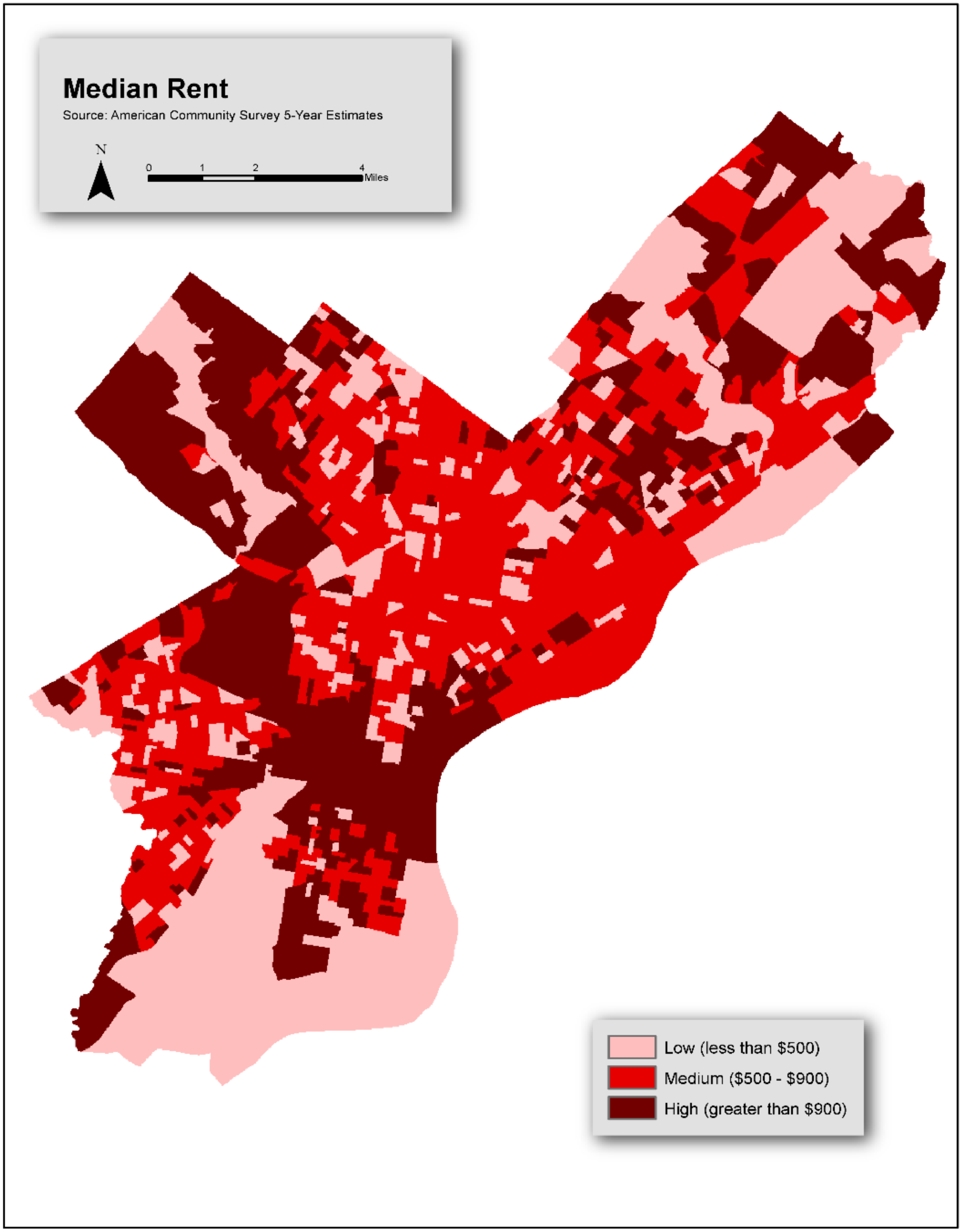


Figure 7. Median Rent from the 2014-2018 American Community Survey

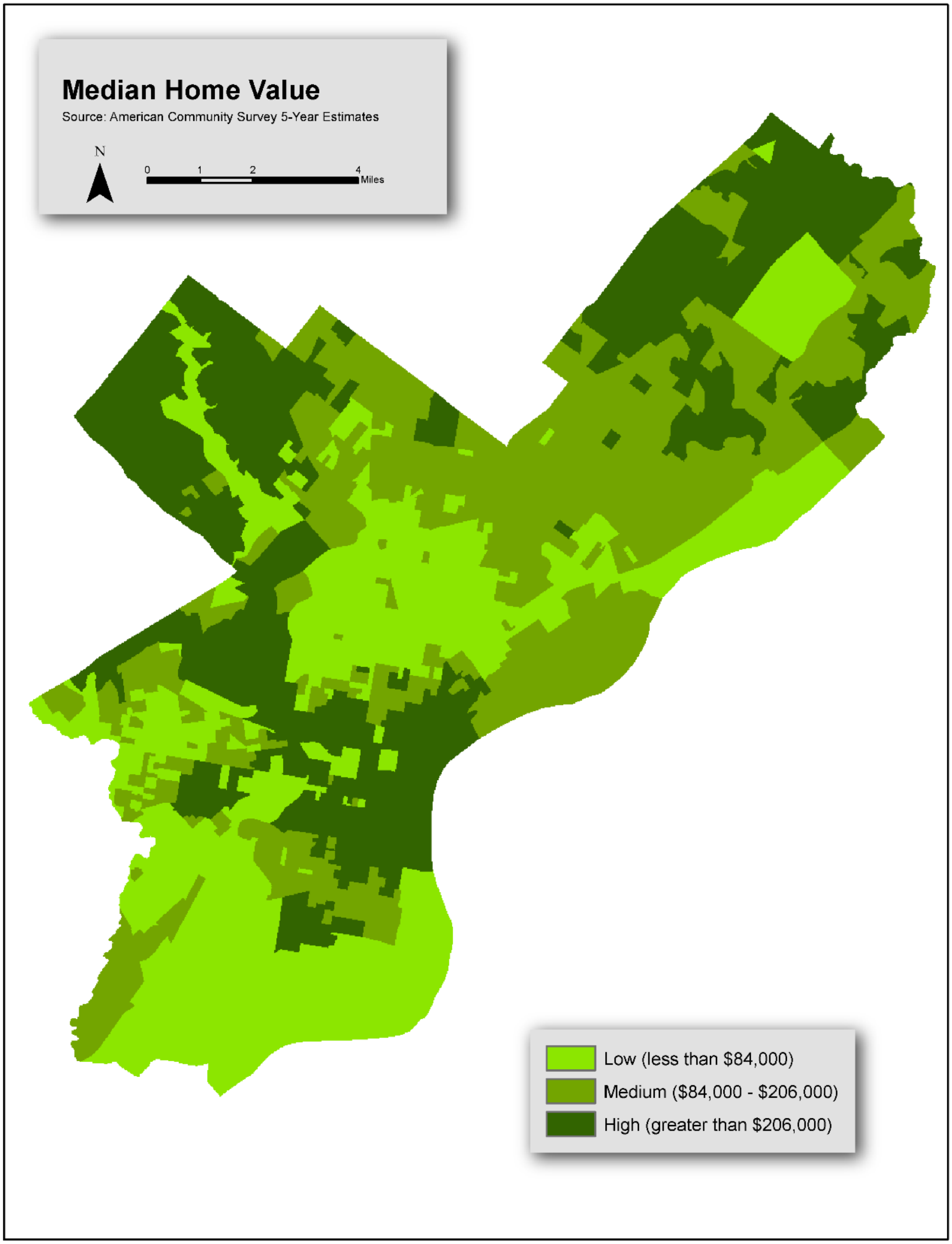


Figure 8. Median Home Value from the 2014-2018 American Community Survey

Correlating Surface Permeability with Demographic Data

I combined the supervised classification with each of the demographic factors using the raster calculator feature in ArcGIS. For this process, I assigned numerical values to both the surface permeability and demographic indicator categorizations. Then I multiplied the values together to create distinct product values.

In order to perform this comparative analysis, both the surface permeability and demographic indicator raster datasets must have the same pixel size.[16] An important issue was then revealed. The Census Bureau aggregates its demographic data by block group, each of which is much larger than 30 m by 30 m. The block-group-based data can easily be apportioned into a 30 m by 30 m raster, but the initial block-group aggregation lowers the overall spatial resolution of this comparative analysis. Therefore, individual-building-level insights are unavailable. Within the scope of this study, this problem could not be worked around. Sufficiently recent single block or residence-based demographic data are not publicly available.[17]

Once distinct categories were created for each correlation, the three maps shown in figures 9, 11, and 13 were created: Surface Permeability Correlated with Median Household Income, Surface Permeability Correlated with Median Rent, and Surface Permeability Correlated with Median Home Value. For each map, I graphed surface areas for the distinct permeability/demographic categories (figures 10, 12, and 14).

---

[16] In this case, the pixel size will be 30 meters (m) x 30 meters (m).
[17] Unaggregated census data becomes public 72 years after its collection.

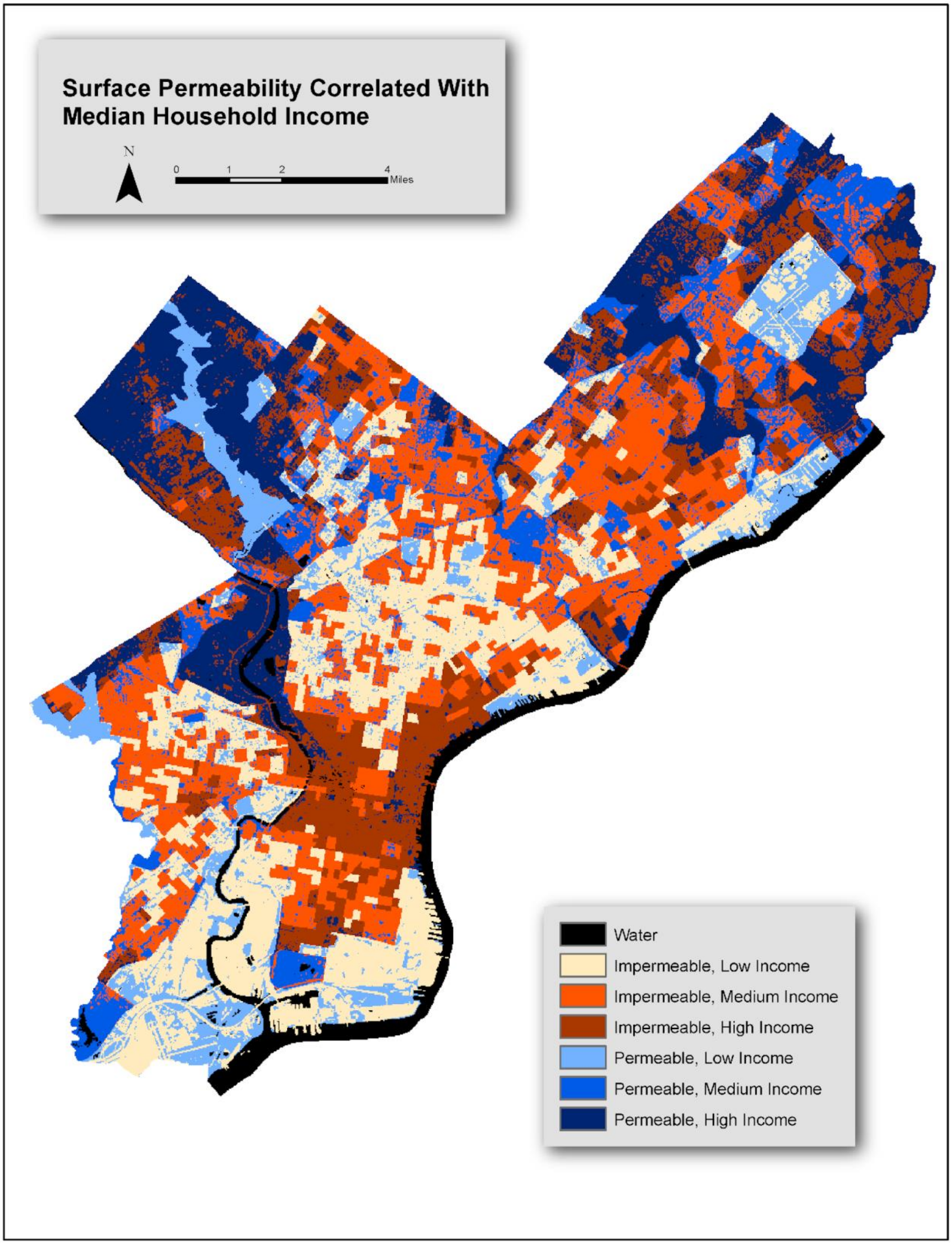


Figure 9. Spatial Overlay Combining Surface Permeability and Household Income

Data

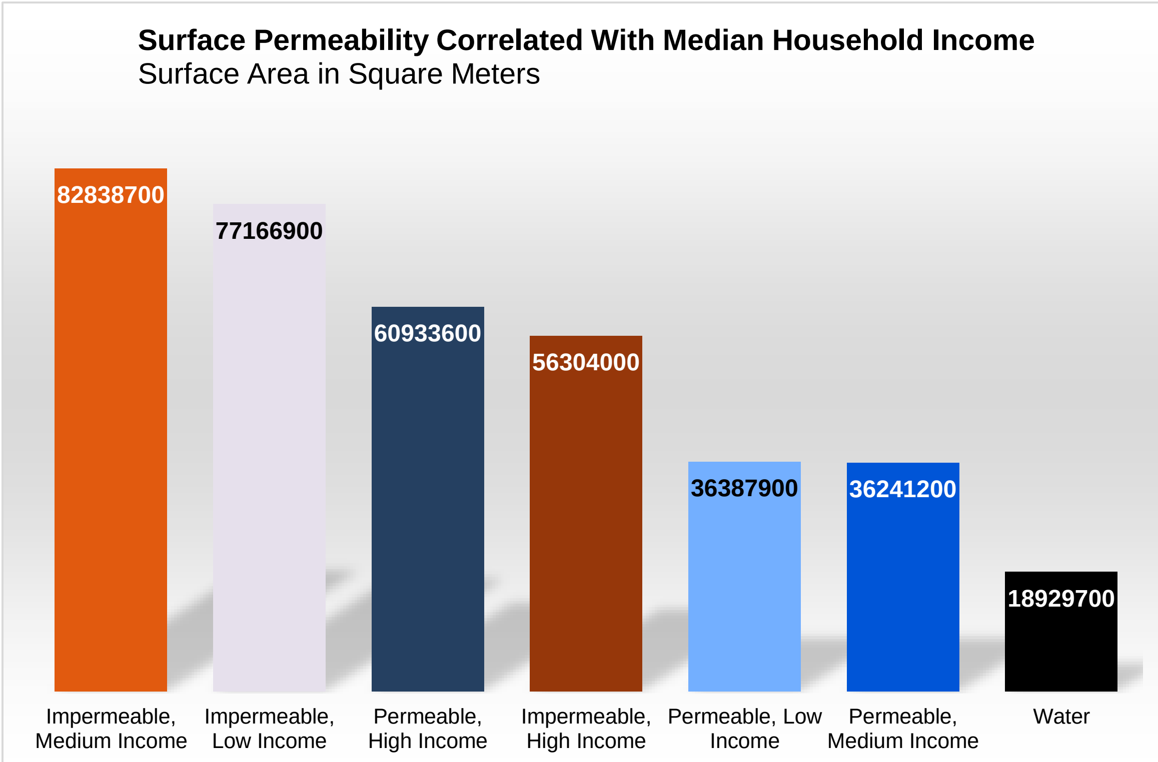


Figure 10. Surface Area Breakdown of Surface Permeability Correlated with Median Household Income

In the correlation between permeability and income, the two most common categories are **Impermeable, Medium Income** and **Impermeable, Low Income**. This result shows the disproportionately large presence of impermeable surfaces in the neighborhoods of less affluent residents. Land falling in these two categories surrounds Center City in a semicircular pattern that extends from south Philadelphia, through west Philadelphia, and up into lower north Philadelphia. The circular pattern is cut off to the east by the Delaware River. **Permeable, High Income** and **Impermeable, High Income** areas form certain distinct patches. The most notable **Impermeable, High Income** patch is Center City. A clear patch of **Permeable, High Income** land can be seen in northwest Philadelphia. Another is observed in the northeast. While income cannot be called a predictor of permeability, a clear correlation exists.

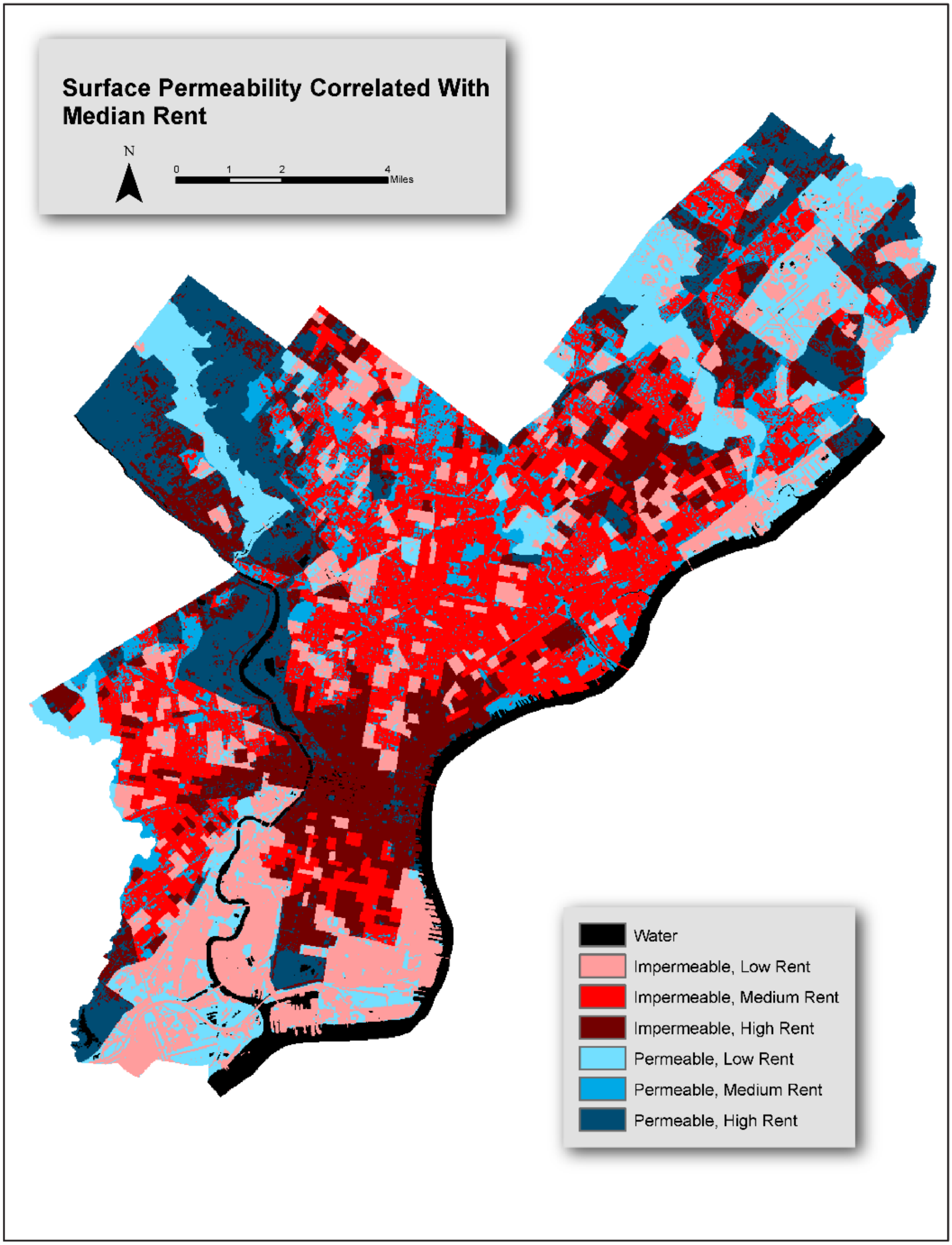


Figure 11. Spatial Overlay Combining Surface Permeability and Rent Data

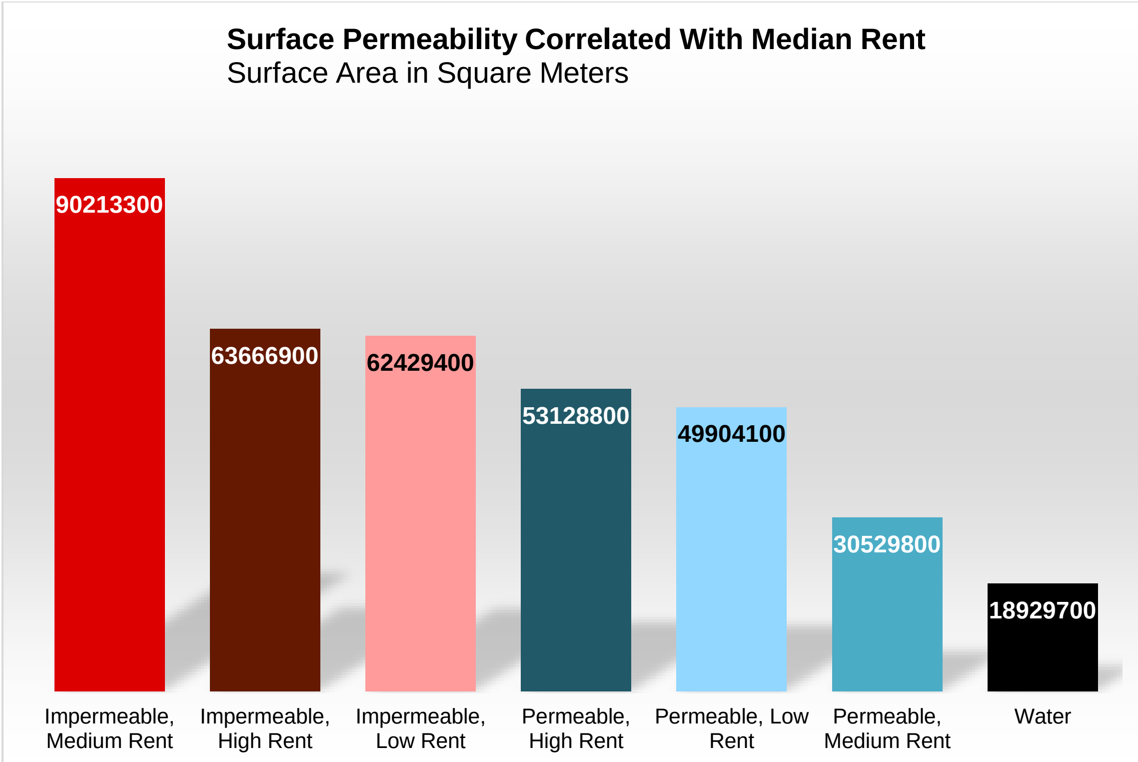


Figure 12. Surface Area Breakdown of Surface Permeability Correlated with Median Rent

The correlation between surface permeability and median rent features numerous distinctly homogenous patches. **Impermeable, Medium Rent** areas dominate in surface area, and clusters of this category can be seen most clearly throughout west and lower north Philadelphia. Patches of **Impermeable, Low Rent** land feature heavily in Philadelphia's southernmost neighborhoods. Center City is predominantly **Impermeable, High Rent**. Additionally, both northwest and northeast Philadelphia show large patches of **Permeable, High Rent** land. **Permeable, Medium Rent** regions were by far the least represented but can be seen dispersed throughout the northernmost regions. As with the correlation between permeability and income, a semicircle of impermeable land that is occupied by less affluent residents (signified by lower rent) is observed: A correlation between permeability and rent certainly exists.

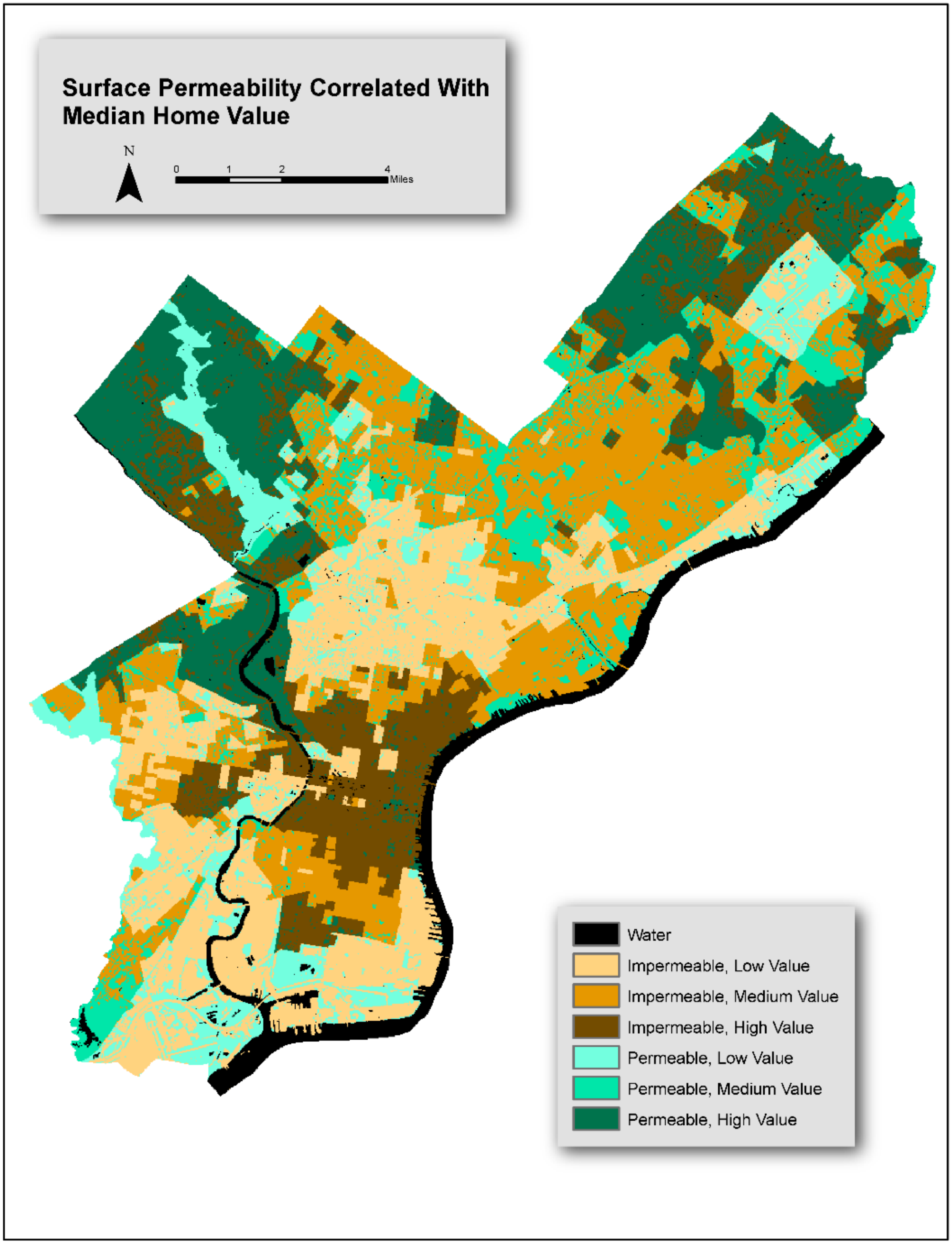


Figure 13. Spatial Overlay Combining Surface Permeability and Home Value Data

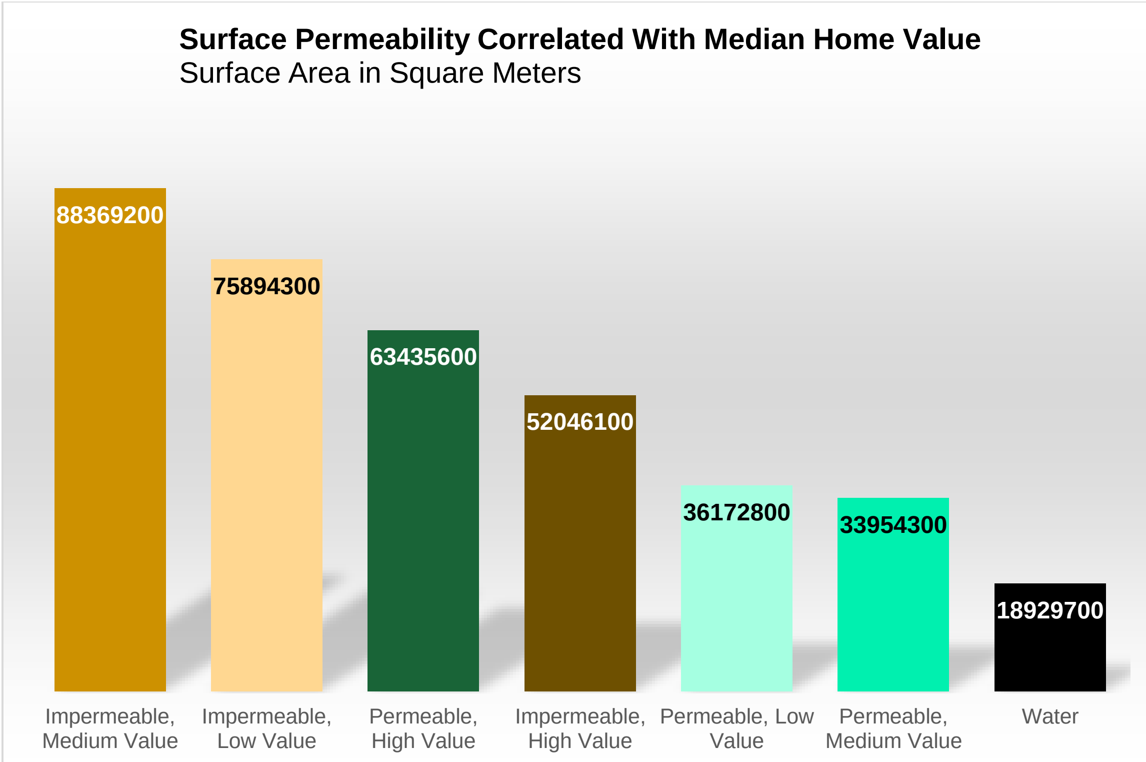


Figure 14. Surface Area Breakdown of Surface Permeability Correlated with Median Home Value

The correlation between surface permeability and median home value shows the largest homogenous patches of all three correlation maps. **Impermeable, Medium Value** regions lead in surface area and are distributed expansively through northern Philadelphia. Another distinct patch can be seen in south Philadelphia. **Impermeable, Low Value** patches form a rough semicircle pattern surrounding central Philadelphia similar to that described in the permeability/income and permeability/rent correlations. **Impermeable, High Value** patches can be easily distinguished in Center City. **Permeable, High Value** patches are clearly present in northwest and northeast Philadelphia. Comparatively less surface area falls in the **Permeable, Low Value** and **Permeable, Medium Value** categories.

Having analyzed the correlation between permeability and three demographic factors, a general trend may be noted: Impermeable surfaces in less-affluent neighborhoods surround an impermeable affluent central district. Permeable affluent neighborhoods are concentrated in northwest and northeast Philadelphia.

Figure 15 shows the sum of each correlation category over all three demographic factors. For example, the **Impermeable, Medium** category is calculated by adding together the **Impermeable, Medium Household Income** and **Impermeable, Medium Rent** and **Impermeable, Medium Home Value** areas. As shown in figure 15 and table 2, high values for the demographic factors are closely split between permeable and impermeable surfaces: 51 percent of the high-demographic-factor areas are permeable. Medium values of the demographic factors possess the largest difference between permeable and impermeable surfaces: 28 percent of medium-demographic-factor areas are permeable. 36 percent of low-demographic-factor areas are permeable. Table 2 shows for the summed surface area of all three demographic indicators, the difference between permeable and impermeable areas. The imbalance toward impermeable surfaces in the medium and low demographic factor categories is clear.

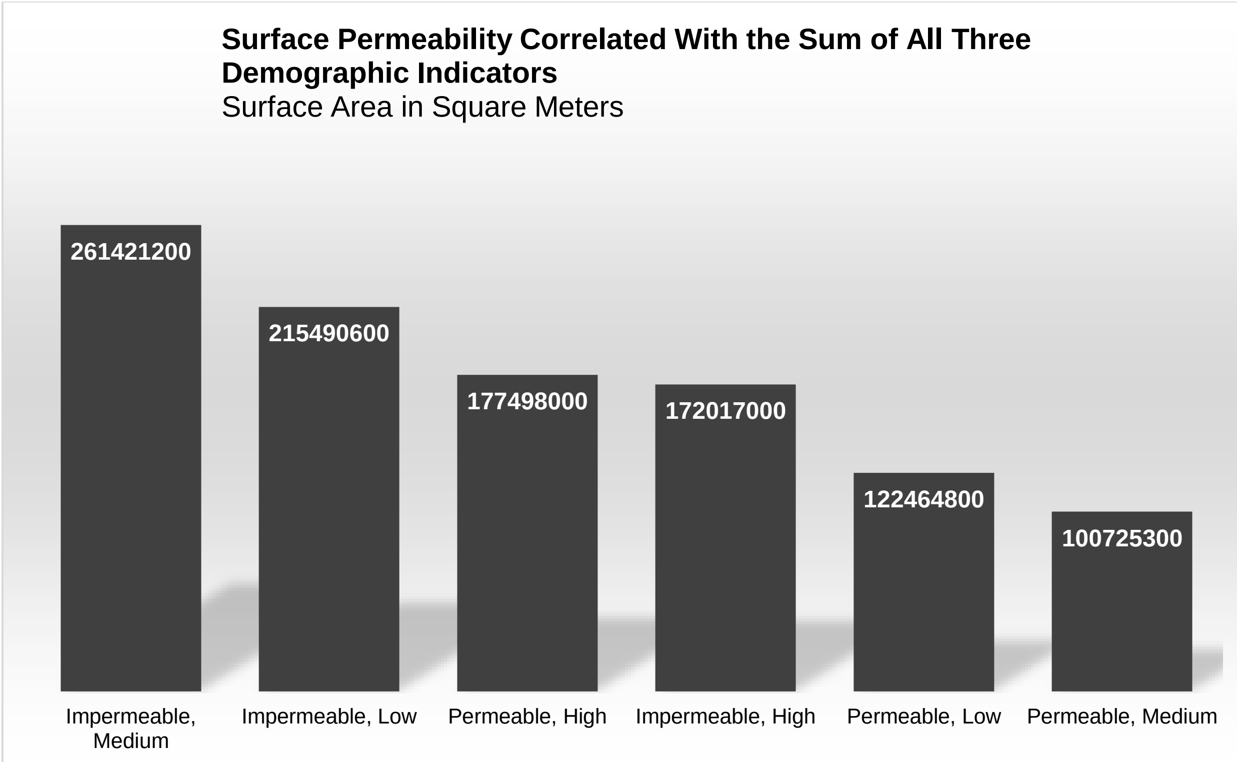


Figure 15. Surface Area Breakdown of Surface Permeability Correlated with All Three Demographic Indicators

Permeable, High minus **Impermeable, High**	5,481,000 m^2
Permeable, Medium minus **Impermeable, Medium**	-160,695,500 m^2
Permeable, Low minus **Impermeable, Low**	-93,025,800 m^2

Table 2. Difference Between Permeable and Impermeable Surface Area over each Demographic Factor Category

Discussion

What does it mean for surface permeability to be equitably distributed across Philadelphia? Should there be a comparable ratio of permeable to impermeable surfaces for each demographic category? There is no "correct" answer to this query. However, the dominance of impermeable surfaces in low and medium demographic factor categories suggests a degree of urban environmental inequity.

This spatial-demographic analysis does not provide a blanket answer as to how surface permeability correlates with the three selected demographic factors in Philadelphia. At the finest level, neighborhood specific assessments are possible. For example, Center City Philadelphia can be characterized as both affluent—in this case defined as majority high income, high rent, and high home value—and covered with mostly impermeable surfaces.

There are several broad themes that can be extracted from the three correlation maps and associated bar graphs and tables. More affluent neighborhoods are concentrated in the center of Philadelphia and in the northwest and northeast corners. In the heart of Philadelphia, land is almost completely impermeable: Center City and the surrounding neighborhoods are densely developed, with only a scattering of public green spaces such as Rittenhouse Square Park. This central region appeals to those who desire high-density living, seek minimal distance to Philadelphia's economic center, and can afford the associated costs of living. On the other hand, in Philadelphia's largely affluent northwest and northeast corners, land is largely permeable. This fact can perhaps be explained by identifying these outer neighborhoods as the transition point between the city and its less-developed suburbs. Wealthy residents of these northern areas still need to be close to

Philadelphia's economic center, yet they also are looking to escape problems associated with high-density urban living. In between these two affluent sections is a semicircle that contains the bulk of lower-income, lower-rent, and lower-home-value communities.

As shown in table 2, Philadelphia's wealthiest communities are about equally divided between permeable and impermeable surfaces. However, this division is not random within wealthy communities: The wealthy central neighborhoods are predominantly impermeable, while the wealthy outlying communities are generally permeable. On the other hand, in medium and low-demographic-factor communities, impermeable surfaces far outnumber permeable surfaces. Generally, it is much likelier to encounter a permeable surface in an affluent community than in a lower income community.

The overall increase in permeability with distance from the city center is reflective of Philadelphia's historical development. The older more central regions were developed in an era when surface permeability was not considered a priority. The variations of socioeconomic status with distance from the center are consistent with Burgess's concentric zone model (Burgess, 1924). Together, these patterns in permeability and demographics yield the correlations observed in this study, in which the relatively affluent choose between the impermeable city center or a permeable outlying zone. Thus, wealthy residents have the option to live in the dense and impermeable Center City or in the permeable neighborhoods bordering Philadelphia's northern suburbs. Less wealthy residents have few alternatives besides living in the impermeable semicircle that envelops Center City.

Is permeability an indicator of the cost of living and neighborhood affordability? What kind of value can be attributed to permeable surfaces? The answers vary from neighborhood to neighborhood. In regions bordering the suburbs, permeable land appears to be highly coveted. Residents of these areas likely have more space and therefore value permeable open green space such as private lawns and backyards. In the more densely developed portions of Philadelphia, permeability is not a clear distinguishing factor between affluent regions such as Center City and the less-affluent southern neighborhoods.

In summary, the most notable patterns that can be identified are as follows. 1) Philadelphia's Center is largely impermeable and affluent. 2) Moving away from the city's center, the land is still largely impermeable but less affluent. 3) Nearing the border between suburbs and city in Philadelphia's northwest and northeast corners, the proportion of permeable and affluent neighborhoods increases.

Conclusion

This research has revealed a spatially and socially uneven distribution of permeable surfaces across Philadelphia. Specifically, permeable surfaces are concentrated in the relatively affluent northwest and northeast corners of the city. While the methods of this study were conducted successfully and the research questions were addressed, there is potential for further research on this topic. Extensions of this work could begin by incorporating more demographic factors into the correlation analysis. As stated in the Demographic Analysis section, household income, rent, and home value do not paint a complete portrait of any community. Demographic factors such as race, ethnicity, land value, and highest attained level of education might be illuminating. Additionally, non-demographic variables such as land use and proximity to Center City could show illuminating correlations with surface permeability.

In the future, the methods of this study can be used to assess surface permeability in other cities besides Philadelphia. For example, as mentioned in the literature review, New York City and Washington D.C. have concerns regarding surface permeability, runoff, and combined sewer systems. At present, this study can help planners and government officials in Philadelphia to assess permeability with high spatial resolution and create narratives relating environmental and demographic data. Supervised learning tasks and correlations with demographics can help identify ideal locations for investments in GI that equitably benefit both affluent and lower income communities.

Peering into the future, this study aligns with Philadelphia's goal to become a smarter city. In 2019, the municipal government released a document called the SmartCityPHL Roadmap, which outlines the applications, objectives, and values of local urban technology solutions. This paper's methodology can be extended to inform green-tech assessments of the environmental and health concerns associated with impermeable surfaces.

If the reader is to come away from this paper with one takeaway, it would be that surface permeability in Philadelphia cannot be described in one blanket statement. As explained in the literature review, permeable surfaces in the form of GI appear in many variations. Surface permeability has a complicated correlation with household income, rent, and home values in Philadelphia. In broadest terms, permeable surfaces are not proportionally distributed among affluent and lower income communities. Despite the complex nature of these findings, the author –and hopefully the reader–have come away from this research with a more holistic and extensive understanding of surface permeability in Philadelphia.

Bibliography

Bauers, S. (2012, April 11). *Phila. embarks on green stormwater management*. Philadelphia Inquirer.

Burgess, E. W. (1924). The growth of the city: An introduction to a research project. *Publications of the American Sociological Society 43*, 85-97.

Dolowitz, D. P., Bell, S., & Keeley, M. (2018). Retrofitting urban drainage infrastructure: Green or grey? *Urban Water Journal 15*, 83-91.

Fitzgerald, J., and Laufer, J. (2017). Governing green stormwater infrastructure: The Philadelphia experience. *Local Environment 22*, 256-268.

Kummer, F. (2019, September 13). *The secret scourge of climate change? More raw sewage in Philadelphia's waterways*. Philadelphia: Philadephia Inquirer.

Labib, S. M., & Harris, A. (2018). The potentials of sentinel-2 and landsat-8 data in green infrastructure extraction, using object based image analysis (OBIA) method. *European Journal of Remote Sensing* 51, 231-240.

Montalto, F., Behr, C., Alfredo, K., Wolf, M., Arye, M., & Walsh, M. (2007). Rapid assessment of the cost-effectiveness of low impact development for CSO control. *Landscape and Urban Planning* 82, 117-131.

Okujeni, A., Canters, F., Cooper, S. D., Degerickx, J., Heiden, U., Hostert, P., Priem, F., Roberts, D. A., Somers, B., van der Linden, S. (2018). Generalizing machine learning regression models using multi-site spectral libraries for mapping vegetation-impervious-soil fractions across multiple cities. *Remote Sensing of Environment* 216, 482-496.

Padmanaban, R., Bhowmik, A. K., Cabral, P. (2019). Satellite image fusion to detect changing surface permeability and emerging urban heat islands in a fast-growing city. *PLoS ONE* 14, e0208949.

Philadelphia, City of. (2014). *Green Streets Design Manual*. http://www.mapc.org/wp-content/uploads/2017/10/GSDM_FINAL_20140211.pdf

Philadelphia, City of. (2019). *SmartCityPHL Roadmap*. https://www.phila.gov/media/20190204121858/SmartCityPHL-Roadmap.pdf

Sansom, G., Berke, P., McDonald, T., Shipp, E., & Horney, J. (2016). Confirming the environmental concerns of community members utilizing participatory-based research in the Houston neighborhood of Manchester. *International Journal of Environmental Research and Public Health* 13, 839.

Taramelli, A., Lissoni, M., Piedelobo, L., Schiavon, E., Valentini, E., Nguyen Xuan, A., & González-Aguilera, D. (2019). Monitoring green infrastructure for natural water retention using Copernicus Global Land Products. *Remote Sensing* 11, 1583.

Weigand, M., Wurm, M., Dech, S., & Taubenböck, H. (2019). Remote sensing in environmental justice research – A review. *Geo-Information* 8, 20.

Xu, X., Schreiber, D. S. P., Lu, Q., & Zhang, Q. (2018). A GIS-based framework creating green stormwater infrastructure inventory relevant to surface transportation planning. *Sustainability* 10, 4710.

Appendix

Script for Supervised Classification of Surface Permeability

Programming Language: R

Integrated Development Environment: RStudio

```
# clear the global environment
rm(list=ls())
# set the working directory
setwd("C:/Users/Classification")

#####
# load libraries
library(raster)
library(tidyverse)
library(sf)
library(rpart)
library(caret)
library(forcats)
library(rpart.plot)
library(rasterVis)
library(mapedit)
library(mapview)
library(magrittr)
library(ggplot2)

#####
# 2020 imagery preparation and processing

# bring in 2020 satellite imagery as individual bands
band1 <- raster("2020_Data/band1.tif")
band2 <- raster("2020_Data/band2.tif")
band3 <- raster("2020_Data/band3.tif")
band4 <- raster("2020_Data/band4.tif")
band5 <- raster("2020_Data/band5.tif")
band6 <- raster("2020_Data/band6.tif")
band7 <- raster("2020_Data/band7.tif")
# band8 <- raster("2020_Data/band8.tif") # band 8 omitted because
of an incompatible extent
band9 <- raster("2020_Data/band9.tif")
band10 <- raster("2020_Data/band10.tif")
band11 <- raster("2020_Data/band11.tif")

# stack the bands into a multi-band raster
image <- stack(band1, band2, band3, band4, band5, band6, band7, #
band8,band9, band10, band11)

# print image properties
nlayers(image)
crs(image)
res(image)

#####
```

```
# create initial plots of 2020 imagery

# plot the true color composite
par(col.axis="white",col.lab="white",tck=0)
plotRGB(image, r = 4, g = 3, b = 2,
        stretch = "lin", main = "2020 True Color Composite")

# export the true color composite to the working directory
trueColor <- stack(band2, band3, band4)
path <- "2020_truecolorcomposite.tif"
writeRaster(trueColor, filename=path)

# plot the false color composite
par(col.axis="white",col.lab="white",tck=0)
plotRGB(image, r = 5, g = 4, b = 3,
        stretch = "lin", main = "2020 False Color Composite")

#####
# preparation for predictive modeling

# read-in training data
training_points <- st_read("2020_Data/training_points.shp")

# extract the spectral values for each training point
training_points <- as(training_points, 'Spatial')
df <- raster::extract(image, training_points) %>%
  round()

# create spectral profiles of each classification category
(permeable, etc.)
profiles <- df %>%
  as.data.frame() %>%
  cbind(., training_points$id) %>%
  rename(id = "training_points$id") %>%
  na.omit() %>%
  group_by(id) %>%
  summarise(band1 = mean(band1),
            band2 = mean(band2),
            band3 = mean(band3),
            band4 = mean(band4),
            band5 = mean(band5),
            band6 = mean(band6),
            band7 = mean(band7),
            # band8 = mean(band8),
            band9 = mean(band9),
            band10 = mean(band10),
            band11 = mean(band11)) %>%
  mutate(id = case_when(id == 1 ~ "permeable",
                        id == 2 ~ "impermeable",
                        id == 3 ~ "water")) %>%
  as.data.frame()
head(profiles)

# plot the spectral profiles of each classification category
across each band
profiles %>%
  select(-id) %>%
```

```
  gather() %>%
  mutate(class = rep(c("Permeable", "Impermeable", "Water"),
10)) %>%
  ggplot(data = ., aes(x = fct_relevel(as.factor(key),
                                        levels = c("band1",
"band2", "band3", "band4","band5", "band6", "band7", # "band8",
"band9", "band10", "band11")), y = value,
                        group=class, color = class)) +
  geom_point(size = 2.5) +
  geom_line(lwd = 1.2) +
  scale_color_manual(values=c("#e8cf7d", "#2a7332", "#0032a0")) +
  labs(title = "Spectral Profiles",
       x = "",
       y = "Surface Reflectance") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_line(color = "gray", size =
0.5),
        panel.grid.minor = element_line(color = "gray", size =
0.5),
        axis.ticks = element_blank())

# create a histogram of spectral profiles
profiles %>%
  select(-id) %>%
  gather() %>%
  mutate(class = rep(c("permeable", "impermeable", "water"),
10)) %>%
  ggplot(., aes(x=value, group=as.factor(class),
fill=as.factor(class))) +
  geom_density(alpha = 0.75) +
  geom_vline(data = . %>% group_by(class) %>% summarise(grp.mean
= mean(value)),
             aes(xintercept=grp.mean, color = class),
linetype="dashed", size=1) +
  scale_fill_manual(values=c("#e8cf7d", "#2a7332", "#0032a0"),
                    name = "class") +
  scale_color_manual(values=c("#e8cf7d", "#2a7332", "#0032a0")) +
  theme(panel.background = element_blank(),
        panel.grid.major = element_line(color = "gray", size =
0.5),
        panel.grid.minor = element_line(color = "gray", size =
0.5),
        axis.ticks = element_blank()) +
  labs(x = "Reflectance Value",
       y = "Density",
       title = "Density Histograms of Spectral Profiles",
       subtitle = "Vertical lines represent mean group
reflectance values")

#####
# predict surface permeability across Philadelphia for 2020

# combine spectral values with their surface permeability class
(permeable, etc.)
df <- data.frame(training_points$id, df)
model.class <- rpart(as.factor(training_points.id)~., data = df,
method = 'class')
```

```
# plot a decision tree
rpart.plot(model.class, box.palette = 0, main = "Classification
Decision Tree", tweak = 1.1)

# predict surface permeability across Philadelphia
pr <- predict(image, model.class, type ='class', progress =
'text') %>%
  ratify()
levels(pr) <- levels(pr)[[1]] %>%
  mutate(legend = c("permeable", "impermeable", "water"))

# plot it
levelplot(pr, maxpixels = 1e6,
          col.regions = c("#2a7332", "#e8cf7d", "#0032a0"),
          scales=list(draw=FALSE),
          main = "Supervised Classification of Imagery")

# export the classification raster
path2 <- "2020_classification.tif"
writeRaster(pr, filename=path2)

#####
# model validation

# compare the training data to the results of the model
test <- raster::extract(pr, training_points) %>%
  as.data.frame() %>%
  rename(id = ".")

testProbs <- data.frame(
  obs = as.factor(training_points$id),
  pred = as.factor(test$id)
) %>%
  mutate(correct = ifelse(obs == pred, 1, 0))

# plot a confusion matrix
confMatrix <- confusionMatrix(testProbs$obs, testProbs$pred)
confMatrix
```