9/10/2019

IST 687 Project

California Protected Land Analysis

Lauren Lawless Dan Kochman David Yamin

Table of Contents

[Introduction 2](#_Toc18322209)

[1. Project Background and Description 2](#_Toc18322210)

[2. Project Scope and Context of this Analysis 2](#_Toc18322211)

[Business Questions 2](#_Toc18322212)

[3. What are the Business Questions? 2](#_Toc18322213)

[Data Acquisition, Cleaning, Transformation, Munging 3](#_Toc18322214)

[4. Describe your data acquisition process 3](#_Toc18322215)

[5. What data did you select? All of it? A subset of it? Why? 3](#_Toc18322216)

[6. What was your initial quality assessment? 3](#_Toc18322217)

[7. What fields/variables did you finally decide on? Why? 3](#_Toc18322218)

[8. Provide a data dictionary 3](#_Toc18322219)

[9. Provide data descriptive statistics, rows, structure, etc. 5](#_Toc18322220)

[10. Did you have to do any cleansing? If so, describe it. 5](#_Toc18322221)

[11. Were there any interesting findings? If so, what were they? 6](#_Toc18322222)

[Descriptive Statistics 6](#_Toc18322223)

[12. Provide demographic statistics 6](#_Toc18322224)

[13. Any early observations, points of interest, interpretation, interesting findings? 9](#_Toc18322225)

[14. Graphs, charts, tables, visuals, text 9](#_Toc18322226)

[Use of Modeling Techniques and Interpretations 9](#_Toc18322227)

[15. Linear regression modeling 18](#_Toc18322228)

[Overall Interpretation of Results/Actionable Insights 48](#_Toc18322229)

[References 48](#_Toc18322230)

[Appendix – RStudio Code 49](#_Toc18322231)

# Introduction

## Project Background and Description

This project is not just an academic exercise of churning through data and analyzing it to answer questions for a class; but to add clarity to a potential problem happening today to real people. Having access to protected land, such as parks, is a quality of life benefit that everyone pays into, but to which not everyone may have equal access. Though it may not be the of utmost concern to most people, if the data shows there is indeed inequality, then it is yet another data point among a growing number of observations that present a country whose rhetoric of freedom and equality are incongruent with its actions.

## Project Scope and Context of this Analysis

The scope of this project encompasses 2010 census data from the state of California. It identifies each county, broken into census tracts, with metrics ranging from household income and housing price to racial and education breakdowns. The ancillary datasets contain geographical location data, as well as voter information of the citizenry. The goal of collecting and analyzing this information is to see if there is inequity among Californians when it comes to accessing protected land. If there is any kind of inequality, we would be looking at possible ways to improve access for currently underserved communities.

# Business Questions

## What are the Business Questions?

1. Can we show inequality in access to protected land areas?
   1. Is there a difference in access for people in rural versus suburban versus urban counties?
      1. Is there a relationship between rural, suburban and urban incomes?
         1. Does that factor into median housing price?
   2. Is there a relationship between race and access (distance\_to\_tract)?
      1. Is there a relationship between race and income (average\_income?)
      2. Is there a relationship between race and housing price (median\_housing\_price)?
   3. Is there a relationship between education and access (distance\_to\_tract)?
      1. Is there a relationship between education and income (average\_income)?
      2. Is there a relationship between education and housing price (median\_housing\_price)?
   4. What are the primary indicators of access, all else constant?
2. If there is inequality, what might reduce it?
   1. Does access increase with voter registration?
   2. Does access increase with voter participation?
   3. Does access vary based on county-wide party affiliation?
      1. If there is a relationship, does it still exist when considering population density classifications (rural/suburban/urban)?
   4. Are there relationships between voter participation/registration and other demographic characteristics?
      1. Race?
      2. Education?
      3. Income?
      4. Housing Price?
3. Can we use a small set of variables to identify non-rural census tracts where voter registration is low and efforts to improve it would be most beneficial to the community?

# Data Acquisition, Cleaning, Transformation, Munging

## Describe your data acquisition process

1. Protected Land - Lauren had a dataset from her undergrad days concerning protect land in California.
2. Voter Registration – Voter registration and participation information.
3. Party Registration – Dave got this data from a 2011 registration report posted by the California Secretary of State.
4. Center – Geographical data for choropleth map creation.

## What data did you select? All of it? A subset of it? Why?

We chose to select all the data. There was a lot of it, but not so much where our computers couldn’t handle it. Since our hardware wasn’t a limiting factor, we opted to not use a subset.

## What was your initial quality assessment?

The quality of the data was pretty good. Several variables had blanks, but we decided, after munging the data, that it is best to leave it in its original state.

## What fields/variables did you finally decide on? Why?

We chose to use all the variables available. Our goal was to see if there were relationships between the fields which could lead to variation in access to protected lands. If there is variation in access, then we would investigate what we could change to increase equity of access for California residents.

## Provide a data dictionary

*(Double-click to enlarge)*



## Provide data descriptive statistics, rows, structure, etc.

The complete dataset of ProtectedLand when joined with other datasets has 5936 observations (rows) and 78 variables (columns). The following shows the structure of the final dataset:

'data.frame': 5936 obs. of 80 variables:

$ gisjoin : Factor w/ 5936 levels "G0600010400100",..: 1 2 3 4 5 6 7 8 9 10 ...

$ county : chr "Alameda" "Alameda" "Alameda" "Alameda" ...

$ year : int 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...

$ tract\_id : int 400100 400200 400300 400400 400500 400600 400700 400800 400900 401000 ...

$ census\_tract\_population : int 2937 1974 4865 3703 3517 1571 4206 3594 2302 5678 ...

$ average\_income : int 117910 81758 50013 49209 34284 29966 23467 30753 28680 19205 ...

$ median\_housing\_price : int 1000001 902400 740000 768000 592100 553700 453500 537600 470500 397700 ...

$ county\_number : int 1 1 1 1 1 1 1 1 1 1 ...

$ tract : int 400100 400200 400300 400400 400500 400600 400700 400800 400900 401000 ...

$ distance\_to\_tract : num 0.1464 0.661 0.2853 0.0942 0.4192 ...

$ county\_population : int 1193529 1193529 1193529 1193529 1193529 1193529 1193529 1193529 1193529 1193529 ...

$ population\_density\_per\_square\_mile : num 2044 2044 2044 2044 2044 ...

$ rural : int 0 0 0 0 0 0 0 0 0 0 ...

$ suburb : int 0 0 0 0 0 0 0 0 0 0 ...

$ urban : int 1 1 1 1 1 1 1 1 1 1 ...

$ share\_white : num 0.708 0.783 0.669 0.655 0.506 ...

$ share\_black : num 0.0477 0.0157 0.1052 0.121 0.2653 ...

$ share\_native\_american : num 0.000681 0.001013 0.001644 0.00135 0.004549 ...

$ share\_asian : num 0.1553 0.074 0.0861 0.0729 0.0591 ...

$ share\_pacific\_islander : num 0.00272 0 0.00103 0.00162 0.00114 ...

$ share\_other : num 0.00851 0.00152 0.00411 0.00567 0.00654 ...

$ share\_2plus : num 0.0378 0.0481 0.0506 0.0532 0.0611 ...

$ share\_hispanic : num 0.0398 0.0765 0.082 0.0897 0.0967 ...

$ pop25plus : int 2276 1570 3723 3134 2768 1209 2787 2051 1699 4088 ...

$ lessthanHS : num 0 0.0159 0.0924 0.0313 0.0372 ...

$ HSdiploma : num 0.0378 0.0783 0.0873 0.0408 0.1322 ...

$ someCollege : num 0.0927 0.0675 0.1091 0.1142 0.2041 ...

$ Associates : num 0.0277 0.028 0.0365 0.0131 0.0744 ...

$ Bachelors : num 0.355 0.342 0.319 0.42 0.339 ...

$ Masters : num 0.246 0.297 0.257 0.22 0.128 ...

$ Professional : num 0.134 0.079 0.0381 0.0788 0.0401 ...

$ Doctorate : num 0.1068 0.0917 0.0612 0.082 0.0448 ...

$ ldist : num -1.922 -0.414 -1.254 -2.363 -0.869 ...

$ ï..Tract : num 6e+09 6e+09 6e+09 6e+09 6e+09 ...

$ Voted : int 1702 1154 2701 2038 1745 738 1703 1289 814 1714 ...

$ Registered : int 2293 1520 3592 2808 2692 1245 2967 2294 1446 3399 ...

$ voting\_age\_pop : int 2510 1639 4105 3131 3067 1347 3577 3022 1880 4519 ...

$ Registered....of.Eligible. : num 0.761 0.761 0.761 0.761 0.761 ...

$ Democratic....of.Eligible. : num 0.568 0.568 0.568 0.568 0.568 ...

$ Republican....of.Eligible. : num 0.157 0.157 0.157 0.157 0.157 ...

$ American.Independent....of.Eligible.: num 0.0186 0.0186 0.0186 0.0186 0.0186 ...

$ Green....of.Eligible. : num 0.0128 0.0128 0.0128 0.0128 0.0128 ...

$ Libertarian....of.Eligible. : num 0.00407 0.00407 0.00407 0.00407 0.00407 ...

$ Peace.and.Freedom....of.Eligible. : num 0.00343 0.00343 0.00343 0.00343 0.00343 ...

$ Other....of.Eligible. : num 0.0153 0.0153 0.0153 0.0153 0.0153 ...

$ Decline.to.State....of.Eligible. : num 0.221 0.221 0.221 0.221 0.221 ...

$ Ratio.Dem.Rep : num 3.62 3.62 3.62 3.62 3.62 ...

$ ï..STATEFP : int 6 6 6 6 6 6 6 6 6 6 ...

$ COUNTYFP : int 1 1 1 1 1 1 1 1 1 1 ...

$ TRACTCE : int 400100 400200 400300 400400 400500 400600 400700 400800 400900 401000 ...

$ POPULATION : int 2937 1974 4865 3703 3517 1571 4206 3594 2302 5678 ...

$ LATITUDE : num 37.9 37.8 37.8 37.8 37.8 ...

$ LONGITUDE : num -122 -122 -122 -122 -122 ...

$ number\_white : num 2078 1546 3256 2424 1778 ...

$ number\_black : num 140 31 512 448 933 ...

$ number\_native\_american : num 2 2 8 5 16 4 14 15 6 20 ...

$ number\_asian : num 456 146 419 270 208 80 249 429 128 418 ...

$ number\_pacific\_islander : num 8 0 5 6 4 0 16 7 5 11 ...

$ number\_other : num 25 3 20 21 23 7 24 7 6 22 ...

$ number\_2plus : num 111 95 246 197 215 68 157 171 132 284 ...

$ number\_hispanic : num 117 151 399 332 340 126 439 367 307 806 ...

$ number\_lessthanHS : num 0 31 450 116 131 ...

$ number\_hsDiploma : num 111 155 425 151 465 190 949 848 241 978 ...

$ number\_someCollege : num 272 133 531 423 718 226 905 487 538 1590 ...

$ number\_Associates : num 81 55 178 48 262 65 297 77 263 349 ...

$ number\_Bachelors : num 1041 675 1550 1556 1193 ...

$ number\_Masters : num 724 587 1249 813 450 ...

$ number\_Professional : num 394 156 186 292 141 87 27 77 39 49 ...

$ number\_Doctorate : num 314 181 298 304 158 12 109 86 0 19 ...

$ number\_Registered : num 1745 1156 2733 2136 2048 ...

$ number\_Democrat : num 1303 864 2041 1595 1529 ...

$ number\_Republican : num 360 239 564 441 423 196 466 360 227 534 ...

$ number\_Independent : num 43 28 67 52 50 23 55 43 27 63 ...

$ number\_Green : num 29 19 46 36 34 16 38 29 19 44 ...

$ number\_Libertarian : num 9 6 15 11 11 5 12 9 6 14 ...

$ number\_PeaceAndFreedom : num 8 5 12 10 9 4 10 8 5 12 ...

$ number\_Party\_Other : num 35 23 55 43 41 19 45 35 22 52 ...

$ number\_Party\_Declined : num 506 335 793 620 594 275 655 506 319 750 ...

$ registration : num 0.914 0.927 0.875 0.897 0.878 ...

$ participation : num 0.678 0.704 0.658 0.651 0.569 ...

## Did you have to do any cleansing? If so, describe it.

We originally used the MICE package to impute NA values from the Protected Land dataset. In doing so, we noticed several models changed between 2% and 6% compared to when we didn’t impute the data. Ultimately, we decided to not impute any of the records. The code has been commented out in the script to show what had been written. It wasn’t necessary to alter any data types for the computations or visualizations we created.

## Were there any interesting findings? If so, what were they?

There weren’t any interesting findings when it comes to data munging. A healthy portion of the records with NAs came from Las Angeles county, with San Diego in a distant second place. Both of those counties are very large, so it isn’t surprising to see NA values associated with them.

# Descriptive Statistics

## Provide demographic statistics

We started off by looking at several demographic makeups to get a better understanding of the data. First and foremost was population distribution:

A picture containing sky, wall, object, flock

Description automatically generated

As you can see, Los Angeles county accounts for ~27% of California’s entire population.

Next, we looked at the population broken down by race:

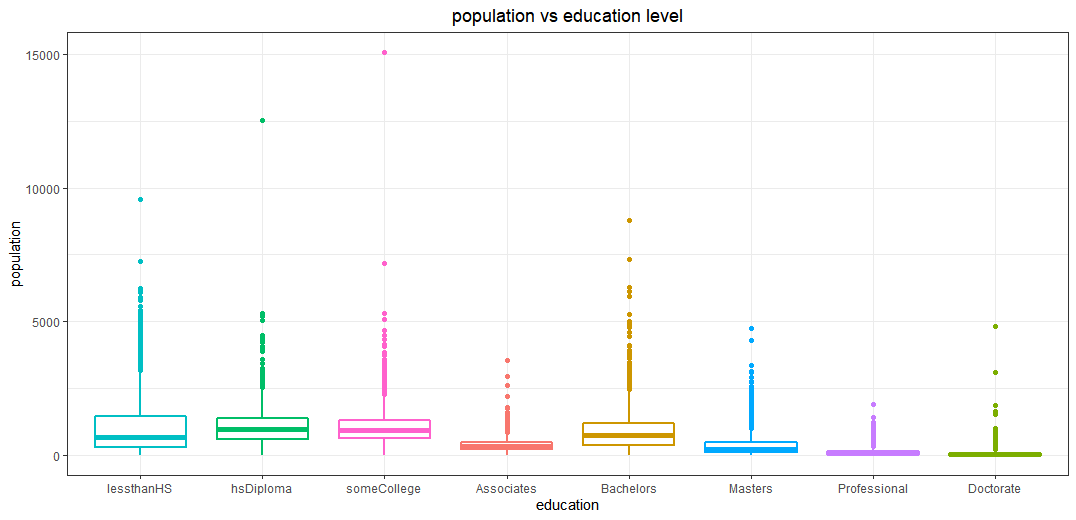
A picture containing text

Description automatically generated

California is primarily made up of white and Hispanic people, with white groups having a slightly

higher median number per census tract, with one notable outlier enclave (Los Angeles).

We then viewed the population through the lens of education:



Educational dispersion is very similar between having a high school diploma and some college. Bachelors degrees and less than high school education levels showed greater dispersion and less centralization around their medians.

Average income was next on our list:

A screenshot of a social media post

Description automatically generated

Most households in California make between $20,000 and $40,000 annually, with the average of the average being just over $30,000. There is a considerable spread of average salaries within the state. To clarify, each census tract had a mean household income value for residents of that census tract. Here, we are displaying the box plot of those averages. This shows us that almost half of the households in California would be considered impoverished (2 parents with 2.5 children has a poverty guideline of $25,750/year per U.S. Department of Health and Human Services).

## Any early observations, points of interest, interpretation, interesting findings?

Los Angeles, Orange, and San Diego counties making up a large portion of the populace for a state the size of California was certainly eye opening. Since population centers tend to have a high cost of living, it was notable to see the average income levels of most households be so low. It was, therefore, not surprising that so many people in California do not have a better education than a high school diploma. With these things in mind, will having more money, a better education, or being part of a racial group grant a person better access (less distance) to protected lands?

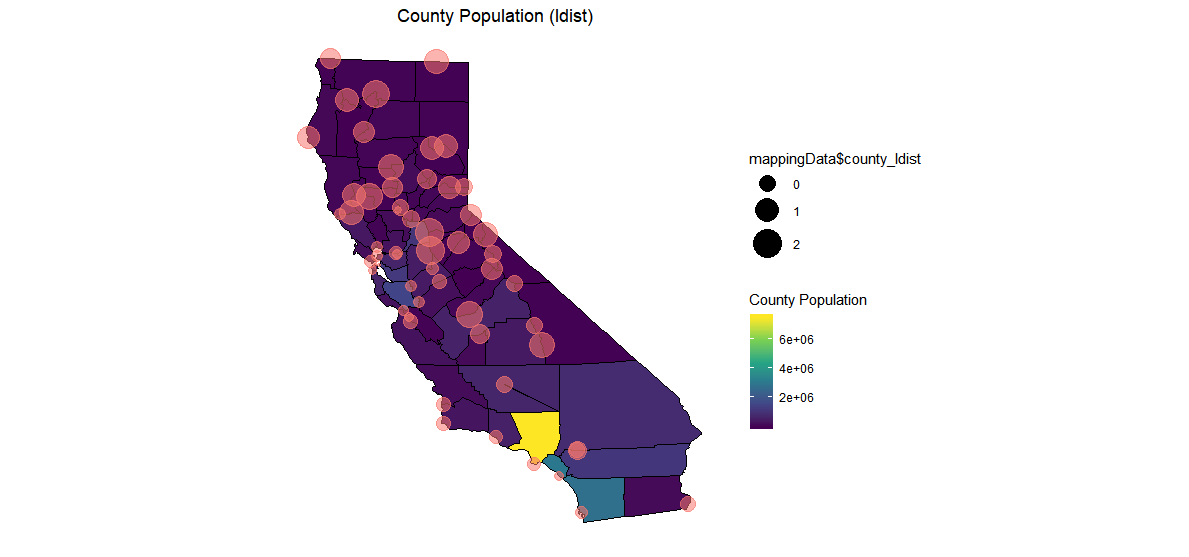
## Graphs, charts, tables, visuals, text

# A screenshot of a cell phone Description automatically generated

A close up of a logo

Description automatically generatedThe above is a histogram showing the distance between a census tract and the nearest protected land area in meters. Since the graph is highly skewed, we used the ldist variable instead:

The following maps of California show distance to protected lands compared to several other variables. The smaller the dot, the closer to protect land the area is.

A screenshot of a cell phone

Description automatically generated

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

A close up of a map

Description automatically generatedA close up of a map

Description automatically generatedA close up of a map

Description automatically generated

# Use of Modeling Techniques and Interpretations

## Linear regression modeling

1. Can we show inequality in access to protected land areas?
   1. Is there a difference in access for people in rural versus suburban versus urban counties?

**lm(ldist ~ urban + suburb, ProtectedLand)**

A screenshot of a cell phone

Description automatically generated

Urban and suburban counties have increased access to protected land areas compared to rural areas. The model has a near-0 p-value and an adjusted R-squared of 11.21%. Suburban county census tracts see a exp(-1.13181)-1 = 67.75% increase in access and urban county census tracts see a 78.48% increase in access.

A close up of a map

Description automatically generated

1ai. Is there a relationship between rural, suburban, and urban incomes?

**lm(average\_income ~ urban + suburb, ProtectedLand)**

**A screenshot of a cell phone

Description automatically generated**

The model is significant and has an adjusted R-squared value of 2.784%. Both coefficients are significant, showing an increase of approximately $10,000 for urban and $4,500 for suburban in average income over rural incomes.

A close up of a map

Description automatically generated

1aia. Does that factor into median housing price?

**lm(median\_housing\_price ~ average\_income + urban + suburb, ProtectedLand)** **A close up of text on a white background

Description automatically generated**

The model is significant and has an adjusted R-squared of 66.23%. Average income and urban locations seem to be the primary drivers of housing prices in California, although suburban housing prices are also much higher.

A close up of a map

Description automatically generated

1b. Is there a relationship between race and access?

**lm(ldist ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLand)**

**A close up of text on a white background

Description automatically generated**

The model is significant and has an adjusted R-squared of 9.437%. Most race/ethnicity coefficients are statistically significant and show slightly increased access compared to white populations. Of notable exception is that Native American populations see highly decreased access (increased distance), which is expected since Native land is not included in the group of designated protected lands.

A bunch of different types of map

Description automatically generated

1bi. Is there a relationship between race and income?

**lm(average\_income ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLand)**

**A close up of text on a white background

Description automatically generated**

The model is significant and has an adjusted R-squared of 54.05%. All minority racial/ethnic populations have lower average incomes compared to white populations except those that identify as other. The coefficients can be interpreted as a 1 percentage point (0.01) increase in the population share of a racial/ethnic minority in a census tract will show a $[coefficient value] change in census tract average income.

A close up of a map

Description automatically generated

1bii. Is there a relationship between race and housing price?

**lm(median\_housing\_price ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLand)**

**A screenshot of text

Description automatically generated**

The model is significant and has an adjusted R-squared of 39.59%. All minority racial/ethnic populations have lower median housing prices compared to white populations except those that identify as Asian or other. The coefficients can be interpreted as a 1 percentage point (0.01) increase in the population share of a racial/ethnic minority in a census tract will show a $[coefficient value] change in census tract median housing price.

A screenshot of a computer

Description automatically generated

1c. Is there a relationship between education and access?

**lm(ldist ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)**

**A screenshot of a cell phone

Description automatically generated**

The model is significant and has an adjusted R-squared of 4.446%. All of the coefficients are positive (decreased access compared to less than High School) except Bachelors and Doctorate, and most are significant except Masters and Doctorate. Notably, professional degree holders have significantly decreased access compared to other education groups.

A screenshot of a computer

Description automatically generated

1ci. Is there a relationship between education and income?

**lm(average\_income ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)**

**A screenshot of a cell phone

Description automatically generated**

The model is significant and has an adjusted R-squared of 78.06%. Most of the coefficients are significant except Associates, with all showing an increase in income (compared to less than High School) except Doctorate.

A close up of a map

Description automatically generated

1cii. Is there a relationship between education and housing price?

**lm(median\_housing\_price ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)**

**A screenshot of text

Description automatically generated**

The model is significant and has an adjusted R-squared of 65.71%. All coefficients are significant except Doctorate, which has a p-value of 0.2902. Compared to less the High School, Bachelors, Masters and Professional degrees see increases in median housing price, while high school diploma, some college, Associates, and Doctorate show decreases in median housing price.

A screenshot of a computer

Description automatically generated

1d. What are primary indicators of access, all else constant?

**lm(ldist ~ average\_income + median\_housing\_price + urban + suburb + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)**

**A screenshot of a cell phone

Description automatically generated**

The model is significant and has an adjusted R-squared of 15.88%. The significant variables are average income, median housing price, urban, suburban, all race/ethnicity variables except other, high school diploma, Bachelors, and Doctorate degrees.

A close up of a map

Description automatically generated

**lm(ldist ~ average\_income + median\_housing\_price + urban + suburb + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLand)**

A close up of text on a white background

Description automatically generated

A close up of a map

Description automatically generated

**lm(ldist ~ average\_income + median\_housing\_price + urban + suburb + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)**

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone on a table

Description automatically generated

What if we only look at urban/suburban areas, where unused green land areas—whether protected or not—are scarcer?

Non-Rural

**lm(ldist ~ average\_income + median\_housing\_price + urban + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

A close up of a map

Description automatically generated

Urban

**lm(ldist ~ average\_income + median\_housing\_price + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandUrban)**

A screenshot of a cell phone

Description automatically generated

A screenshot of a computer

Description automatically generated

Suburban

**lm(ldist ~ average\_income + median\_housing\_price + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandSuburban)**

**A screenshot of a cell phone

Description automatically generated**

A close up of a map

Description automatically generated

Rural

**lm(ldist ~ average\_income + median\_housing\_price + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandRural)** **A screenshot of a cell phone

Description automatically generated**

**A close up of a map

Description automatically generated**

Looking at non-rural areas (urban and suburban combined), we see that a combination of median housing price and average income, when taken with appropriate magnitudes, would oftentimes indicate increased access for higher-income communities (overall negative effect on distance) over lower-income communities. We also see mostly negative coefficients (decreased distance and thus increased access) for more educated communities, compared to mostly positive coefficients (increased distance and thus decreased access) for less educated communities. Therefore, while we cannot establish racial inequality in access to protected lands in urban and suburban areas, we can reasonably state that there is income and education-related inequality.

1. If there is inequality, what might reduce it?

Since protected land areas are commonly established through local and state government groups, departments, or officials, we will investigate the relationship between voter participation, party affiliation, and protected land areas.

2a. Does access increase with voter registration?

**lm(ldist ~ registration, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

The coefficient indicates that in non-rural California, as voter registration (proportion of voting age population that is registered to vote) increases, distance to protected land decreases (increased access).

A screenshot of a map

Description automatically generated

2b. Does access increase with voter participation?

**lm(ldist ~ participation, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

This coefficient indicates that in non-rural California, as voter participation (proportion of voting age population that voted in the last election) increases, distance to protected land also increases (decreased access).

2bi. **lm(ldist ~ registration + participation, ProtectedLandNonRural)**

A screenshot of a cell phone

Description automatically generatedA close up of a map

Description automatically generated

This model shows the same coefficient signs from previous models at high levels of significance, indicating that higher voter registration and lower voter participation would be associated with increased access to protected land areas. Speculatively, this may indicate that voter engagement (interacting with local officials and politics in ways other than voting) may impact local protected areas more so than just voter participation.

2c. Does access depend on county-wide party affiliation?

**lm(ldist ~ Ratio.Dem.Rep, ProtectedLand)**

**A screenshot of a cell phone

Description automatically generated**

The variable Ration.Dem.Rep represents the ratio of the proportion of registered Democrats to the proportion of registered Republicans at the county level. This model indicates that as the ratio increases by county (more registered Democrats compared to registered Republicans), the distance to protected land decreases for census tracts in that county.

A screenshot of a cell phone

Description automatically generated

2ci. If there is a relationship, does it still exist when considering population density classification (rural, suburb, urban)?

**lm(ldist ~ Ratio.Dem.Rep, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

**lm(ldist ~ Ratio.Dem.Rep, ProtectedLandUrban)**

**A screenshot of a cell phone

Description automatically generated**

**lm(ldist ~ Ratio.Dem.Rep, ProtectedLandSuburban)**

**A screenshot of a cell phone

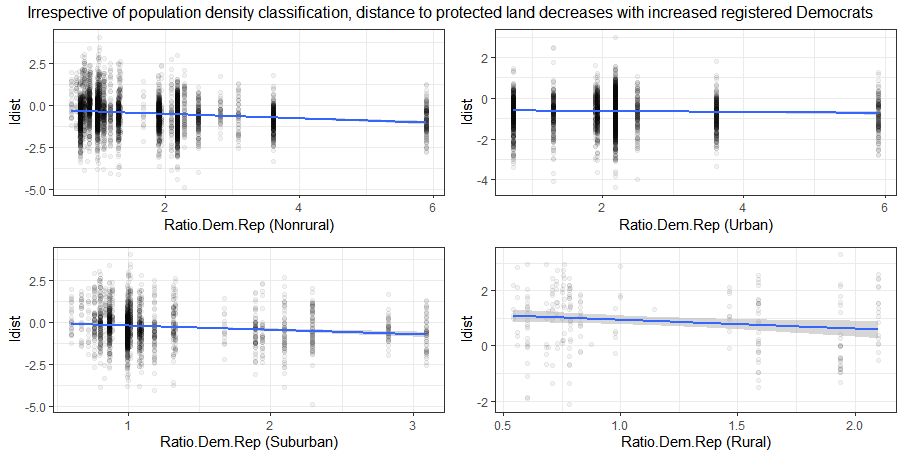
Description automatically generated**

**lm(ldist ~ Ratio.Dem.Rep, ProtectedLandRural)**

**A screenshot of a cell phone

Description automatically generated**

This relationship persists through all population density segments we have examined in this data set.



2d. Are there relationships between voter participation/registration and other demographic characteristics?

2di. Race/ethnicity?

All minority populations show decreased voter registration and decreased voter participation compared to white populations in non-rural counties of California.

**lm(registration ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

A close up of a map

Description automatically generated

**lm(participation ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

A close up of a map

Description automatically generated

2dii. Education?

As education level increases, so does voter registration and voter turnout, except for communities with higher proportions of Doctorates, where we see turnout and participation levels much lower than all other education levels.

**lm(registration ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandNonRural)**

A screenshot of a cell phone

Description automatically generated

A close up of a map

Description automatically generated

**lm(participation ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

**A close up of a map

Description automatically generated**

2diii. Income?

As income increases, so does voter registration and voter participation.

**lm(registration ~ average\_income, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

**A close up of a map

Description automatically generated**

**lm(participation ~ average\_income, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

**A close up of a map

Description automatically generated**

2div. Housing price?

As housing price increases, so does voter registration and voter participation.

**lm(registration ~ median\_housing\_price, ProtectedLandNonRural)** A screenshot of a cell phone

Description automatically generated

**A screenshot of a map

Description automatically generated**

**lm(participation ~ median\_housing\_price, ProtectedLandNonRural)**

**A screenshot of a cell phone

Description automatically generated**

**A screenshot of a map

Description automatically generated**

1. Can we use a small set of variables to identify non-rural census tracts where voter registration is low and efforts to improve it would be most beneficial to the community?

**ksvm(lowReg ~ median\_housing\_price + average\_income + lessthanHS + Ratio.Dem.Rep, data = trainData, kernel = "rbfdot", kpar = "automatic", C = 5, cross = 3, prob.model = TRUE)**

**A screenshot of a cell phone

Description automatically generated**

A close up of a map

Description automatically generated

The Support Vector Machine model correctly identifies 84.97% of the test data observations as “low” or “high” (below or above the mean) voter registration census tracts, based on data for housing price, income, less than high school proportion, and ratio of registered Democrats to Republicans. The model primarily struggles to correctly identify census tracts where the average income is near the mean.

# Overall Interpretation of Results/Actionable Insights

Voter registration, a significant indicator of access to protected land areas by census tract, is reduced in certain communities of color, less educated communities, and lower income communities. Voter participation is similarly impacted in disadvantaged communities, but we do not see increased voter participation have the same association with access. With this knowledge, and without direct intervention to designate protected land areas in disadvantaged communities, we would suggest focusing on increasing voter registration in those communities as a gateway for increased voter engagement. That being said, direct intervention may be necessary, as disadvantaged communities often have more difficulty engaging with politics (lack of flexible working hours, inaccessibility to engagement opportunities).

# References

Census Data - <https://www.census.gov/geographies/reference-files/2010/geo/2010-centers-population.html>

Protected Land Data - <https://www.calands.org/cpad/>

Party Registration Data - <https://www.sos.ca.gov/elections/report-registration/ror-odd-year-11/>

Voting Data - <https://data.chhs.ca.gov/dataset/voter-registration-2002-2010/resource/c384c86a-49d2-4128-8389-b2701ff0bc35>

US HHS - <https://www.federalregister.gov/documents/2019/02/01/2019-00621/annual-update-of-the-hhs-poverty-guidelines>

# Appendix – RStudio Code

#Index - Double click on the bracketed string and press Ctrl+F and then Enter to jump to that section.

# Load and prep [i001]

# Data Munging [i002]

# Descriptive Analysis [i003]

# Descriptive Charts [i003a]

# Descriptive Maps [i003b]

# Advanced Analysis [i004]

#--------------------------------------Load and prep [i001]--------------------------------------------------------------

#Activate libraries.

require(arulesViz)

require(caret)

require(crayon)

require(dplyr)

require(e1071)

require(GGally)

require(ggcorrplot)

require(ggiraph)

require(ggiraphExtra)

require(ggplot2)

require(ggpubr)

require(gridExtra)

require(kernlab)

require(mapproj)

require(maps)

require(mice)

require(readxl)

require(reshape2)

require(VIM)

#Load the Protected Land Data & Voter Data .csv file.

# ProtectedLand <- read.csv("C:\\Dan\\SU\\IST 687 (70869) - Intro to Data Science\\Project\\Protected Land.csv")

# Voting <- read.csv("C:\\Dan\\SU\\IST 687 (70869) - Intro to Data Science\\Project\\Voting.csv")

# Party <- read.csv("C:\\Dan\\SU\\IST 687 (70869) - Intro to Data Science\\Project\\Party.csv")

# Centers <- read.csv("C:\\Dan\\SU\\IST 687 (70869) - Intro to Data Science\\Project\\Centers.csv")

ProtectedLand <- read.csv(file.choose())

Voting <- read.csv(file.choose())

Party <- read.csv(file.choose())

Centers <- read.csv(file.choose())

mappingData <- read\_excel(file.choose())

#--------------------------------------Data Munging [i002]--------------------------------------------------------------

#Clean Data - Rename columns to make them more understandable.

names(ProtectedLand)[2] <- "year"

names(ProtectedLand)[3] <- "county"

names(ProtectedLand)[4] <- "tract\_id"

names(ProtectedLand)[5] <- "census\_tract\_population"

names(ProtectedLand)[6] <- "average\_income"

names(ProtectedLand)[7] <- "median\_housing\_price"

names(ProtectedLand)[8] <- "county\_number"

names(ProtectedLand)[10] <- "distance\_to\_tract"

names(ProtectedLand)[11] <- "county\_population"

names(ProtectedLand)[12] <- "population\_density\_per\_square\_mile"

names(ProtectedLand)[18] <- "share\_native\_american"

names(ProtectedLand)[20] <- "share\_pacific\_islander"

names(Party)[1] <- "county"

names(Voting)[5] <- "voting\_age\_pop"

#Clean Data - Remove " County" from the county field. It's redundant and takes up space on visualizations.

# gsub() replaces the 1st arguement with the 2nd arguement while looking through the 3rd arguement.

ProtectedLand$county <- gsub(" County", "", ProtectedLand$county)

#Merge Party Data

ProtectedLand <- merge(ProtectedLand, Voting, by = "gisjoin")

ProtectedLand <- merge(ProtectedLand, Party, by.x = "county", by.y = "county")

ProtectedLand <- merge(ProtectedLand, Centers, by = "gisjoin")

# ####Using R Mice Package to impute missing values [average\_income column]####

#

# ###Step 1: Confirm the existence of 'NA's within a column

#

# any(is.na(ProtectedLand$average\_income)) ###Returns 'TRUE' Specifically, rows 1918, 2234, 2248, 2819, 4103 and 4213 are 'NA's.

#

# ###Step 2: Create a data.frame of at least two columns to work with

#

# miceAverage\_Income <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$average\_income)

#

# #####Step 3: Probably redundant of Step 1 above, but interesting nonetheless. This function creates a 'pattern' graphic

# ######that illustrates the number of rows in our dummy df with 0 missing values and the rows with 1 missing values.

#

# md.pattern(miceAverage\_Income)

#

# #####Step 4: The R mice function develops potential imputed values for 'NA's using various types of regression.

# imputed\_miceAverage\_Income <- mice(miceAverage\_Income, m=5, method = 'pmm', seed = 101)

#

# ######Step 5: We use the R Mice's 'compete' function to choose the results regression method (here, method '3')

# ######to be imputed in place of the 'NA's

#

# miceAverage\_Income <- complete(imputed\_miceAverage\_Income,3)

#

# ###### Step 6: We replace the old average\_income column values with the column that contains the six imputed values to replace 'NA's

# ProtectedLand$average\_income <- miceAverage\_Income$ProtectedLand.average\_income

#

# #####Step 7: We confirm the absence of 'NA's within the replacement column

#

# any(is.na(ProtectedLand$average\_income))

# ProtectedLand$average\_income [1918] ##### For example, we see that the 'NA' formerly in row 1918 has been replaced with###

# ### with '31418'

# #Clearing unneeded object

# rm(imputed\_miceAverage\_Income)

# rm(miceAverage\_Income)

#

# ######Using R Mice Package to impute missing values [median\_housing\_price]

#

# any(is.na(ProtectedLand$median\_housing\_price))

#

# miceMedianHousingPrice <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$median\_housing\_price)

# md.pattern(miceMedianHousingPrice)

# imputed\_miceMedianHousingPrice <- mice(miceMedianHousingPrice, m=5, method = 'pmm', seed = 101)

# miceMedianHousingPrice <- complete(imputed\_miceMedianHousingPrice,3)

# ProtectedLand$median\_housing\_price <- miceMedianHousingPrice$ProtectedLand.median\_housing\_price

#

# any(is.na(ProtectedLand$median\_housing\_price))

#

# #Clearing unneeded object

# rm(imputed\_miceMedianHousingPrice)

# rm(miceMedianHousingPrice)

#

# ######Using R Mice Package to impute missing values [share\_white]

#

# any(is.na(ProtectedLand$share\_white))

#

# miceshare\_white <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$share\_white)

# md.pattern(miceshare\_white)

# imputed\_miceshare\_white <- mice(miceshare\_white, m=5, method = 'pmm', seed = 101)

# miceshare\_white <- complete(imputed\_miceshare\_white,3)

# ProtectedLand$share\_white <- miceshare\_white$ProtectedLand.share\_white

#

# any(is.na(ProtectedLand$share\_white))

#

# #Clearing unneeded object

# rm(imputed\_miceshare\_white)

# rm(miceshare\_white)

#

# ######Using R Mice Package to impute missing values [share\_black]

#

# any(is.na(ProtectedLand$share\_black))

#

# miceshare\_black <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$share\_black)

# md.pattern(miceshare\_black)

# imputed\_miceshare\_black <- mice(miceshare\_black, m=5, method = 'pmm', seed = 101)

# miceshare\_black <- complete(imputed\_miceshare\_black,3)

# ProtectedLand$share\_black <- miceshare\_black$ProtectedLand.share\_black

#

# any(is.na(ProtectedLand$share\_black))

#

# #Clearing unneeded object

# rm(imputed\_miceshare\_black)

# rm(miceshare\_black)

#

# ######Using R Mice Package to impute missing values [share\_native\_american]

#

# any(is.na(ProtectedLand$share\_native\_american))

#

# miceshare\_native\_american <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$share\_native\_american)

# md.pattern(miceshare\_native\_american)

# imputed\_miceshare\_native\_american <- mice(miceshare\_native\_american, m=5, method = 'pmm', seed = 101)

# miceshare\_native\_american <- complete(imputed\_miceshare\_native\_american,3)

# ProtectedLand$share\_native\_american <- miceshare\_native\_american$ProtectedLand.share\_native\_american

#

# any(is.na(ProtectedLand$share\_native\_american))

#

# #Clearing unneeded object

# rm(imputed\_miceshare\_native\_american)

# rm(miceshare\_native\_american)

#

# ######Using R Mice Package to impute missing values [share\_asian]

#

# any(is.na(ProtectedLand$share\_asian))

#

# miceshare\_asian <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$share\_asian)

# md.pattern(miceshare\_asian)

# imputed\_miceshare\_asian <- mice(miceshare\_asian, m=5, method = 'pmm', seed = 101)

# miceshare\_asian <- complete(imputed\_miceshare\_asian,3)

# ProtectedLand$share\_asian <- miceshare\_asian$ProtectedLand.share\_asian

#

# any(is.na(ProtectedLand$share\_asian))

#

# #Clearing unneeded object

# rm(imputed\_miceshare\_asian)

# rm(miceshare\_asian)

#

# ######Using R Mice Package to impute missing values [share\_pacific\_islander]

#

# any(is.na(ProtectedLand$share\_pacific\_islander))

#

# miceshare\_pacific\_islander <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$share\_pacific\_islander)

# md.pattern(miceshare\_pacific\_islander)

# imputed\_miceshare\_pacific\_islander <- mice(miceshare\_pacific\_islander, m=5, method = 'pmm', seed = 101)

# miceshare\_pacific\_islander <- complete(imputed\_miceshare\_pacific\_islander,3)

# ProtectedLand$share\_pacific\_islander <- miceshare\_pacific\_islander$ProtectedLand.share\_pacific\_islander

#

# any(is.na(ProtectedLand$share\_pacific\_islander))

#

# #Clearing unneeded object

# rm(imputed\_miceshare\_pacific\_islander)

# rm(miceshare\_pacific\_islander)

#

# ######Using R Mice Package to impute missing values [share\_other]

#

# any(is.na(ProtectedLand$share\_other))

#

# miceshare\_other <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$share\_other)

# md.pattern(miceshare\_other)

# imputed\_miceshare\_other <- mice(miceshare\_other, m=5, method = 'pmm', seed = 101)

# miceshare\_other <- complete(imputed\_miceshare\_other,3)

# ProtectedLand$share\_other <- miceshare\_other$ProtectedLand.share\_other

#

# any(is.na(ProtectedLand$share\_other))

#

# #Clearing unneeded object

# rm(imputed\_miceshare\_other)

# rm(miceshare\_other)

#

# ######Using R Mice Package to impute missing values [share\_2plus]

#

# any(is.na(ProtectedLand$share\_2plus))

#

# miceshare\_2plus <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$share\_2plus)

# md.pattern(miceshare\_2plus)

# imputed\_miceshare\_2plus <- mice(miceshare\_2plus, m=5, method = 'pmm', seed = 101)

# miceshare\_2plus <- complete(imputed\_miceshare\_2plus,3)

# ProtectedLand$share\_2plus <- miceshare\_2plus$ProtectedLand.share\_2plus

#

# any(is.na(ProtectedLand$share\_2plus))

#

# #Clearing unneeded object

# rm(imputed\_miceshare\_2plus)

# rm(miceshare\_2plus)

#

# ######Using R Mice Package to impute missing values [share\_hispanic]

#

# any(is.na(ProtectedLand$share\_hispanic))

#

# miceshare\_hispanic <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$share\_hispanic)

# md.pattern(miceshare\_hispanic)

# imputed\_miceshare\_hispanic <- mice(miceshare\_hispanic, m=5, method = 'pmm', seed = 101)

# miceshare\_hispanic <- complete(imputed\_miceshare\_hispanic,3)

# ProtectedLand$share\_hispanic <- miceshare\_hispanic$ProtectedLand.share\_hispanic

#

# any(is.na(ProtectedLand$share\_hispanic))

#

# #Clearing unneeded object

# rm(imputed\_miceshare\_hispanic)

# rm(miceshare\_hispanic)

#

# ######Using R Mice Package to impute missing values [lessthanHS]

#

# any(is.na(ProtectedLand$lessthanHS))

#

# micelessthanHS <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$lessthanHS)

# md.pattern(micelessthanHS)

# imputed\_micelessthanHS <- mice(micelessthanHS, m=5, method = 'pmm', seed = 101)

# micelessthanHS <- complete(imputed\_micelessthanHS,3)

# ProtectedLand$lessthanHS <- micelessthanHS$ProtectedLand.lessthanHS

#

# any(is.na(ProtectedLand$lessthanHS))

#

# #Clearing unneeded object

# rm(imputed\_micelessthanHS)

# rm(micelessthanHS)

#

# ######Using R Mice Package to impute missing values [HSdiploma]

#

# any(is.na(ProtectedLand$HSdiploma))

#

# miceHSdiploma <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$HSdiploma)

# md.pattern(miceHSdiploma)

# imputed\_miceHSdiploma <- mice(miceHSdiploma, m=5, method = 'pmm', seed = 101)

# miceHSdiploma <- complete(imputed\_miceHSdiploma,3)

# ProtectedLand$HSdiploma <- miceHSdiploma$ProtectedLand.HSdiploma

#

# any(is.na(ProtectedLand$HSdiploma))

#

# #Clearing unneeded object

# rm(imputed\_miceHSdiploma)

# rm(miceHSdiploma)

#

# ######Using R Mice Package to impute missing values [someCollege]

#

# any(is.na(ProtectedLand$someCollege))

#

# micesomeCollege <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$someCollege)

# md.pattern(micesomeCollege)

# imputed\_someCollege <- mice(micesomeCollege, m=5, method = 'pmm', seed = 101)

# micesomeCollege <- complete(imputed\_someCollege,3)

# ProtectedLand$someCollege <- micesomeCollege$ProtectedLand.someCollege

#

# any(is.na(ProtectedLand$someCollege))

#

# #Clearing unneeded object

# rm(imputed\_someCollege)

# rm(micesomeCollege)

#

# ######Using R Mice Package to impute missing values [Associates]

#

# any(is.na(ProtectedLand$Associates))

#

# miceAssociates <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$Associates)

# md.pattern(miceAssociates)

# imputed\_Associates <- mice(miceAssociates, m=5, method = 'pmm', seed = 101)

# miceAssociates <- complete(imputed\_Associates,3)

# ProtectedLand$Associates <- miceAssociates$ProtectedLand.Associates

#

# any(is.na(ProtectedLand$Associates))

#

# #Clearing unneeded object

# rm(imputed\_Associates)

# rm(miceAssociates)

#

# ######Using R Mice Package to impute missing values [Bachelors]

#

# any(is.na(ProtectedLand$Bachelors))

#

# miceBachelors <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$Bachelors)

# md.pattern(miceBachelors)

# imputed\_Bachelors <- mice(miceBachelors, m=5, method = 'pmm', seed = 101)

# miceBachelors <- complete(imputed\_Bachelors,3)

# ProtectedLand$Bachelors <- miceBachelors$ProtectedLand.Bachelors

#

# any(is.na(ProtectedLand$Bachelors))

#

# #Clearing unneeded object

# rm(imputed\_Bachelors)

# rm(miceBachelors)

#

# ######Using R Mice Package to impute missing values [Masters]

#

# any(is.na(ProtectedLand$Masters))

#

# miceMasters <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$Masters)

# md.pattern(miceMasters)

# imputed\_Masters <- mice(miceMasters, m=5, method = 'pmm', seed = 101)

# miceMasters <- complete(imputed\_Masters,3)

# ProtectedLand$Masters <- miceMasters$ProtectedLand.Masters

#

# any(is.na(ProtectedLand$Masters))

#

# #Clearing unneeded object

# rm(imputed\_Masters)

# rm(miceMasters)

#

# ######Using R Mice Package to impute missing values [Professional]

#

# any(is.na(ProtectedLand$Professional))

#

# miceProfessional <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$Professional)

# md.pattern(miceProfessional)

# imputed\_Professional <- mice(miceProfessional, m=5, method = 'pmm', seed = 101)

# miceProfessional <- complete(imputed\_Professional,3)

# ProtectedLand$Professional <- miceProfessional$ProtectedLand.Professional

#

# any(is.na(ProtectedLand$Professional))

#

# #Clearing unneeded object

# rm(imputed\_Professional)

# rm(miceProfessional)

#

# ######Using R Mice Package to impute missing values [Doctorate]

#

# any(is.na(ProtectedLand$Doctorate))

#

# miceDoctorate <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$Doctorate)

# md.pattern(miceDoctorate)

# imputed\_Doctorate <- mice(miceDoctorate, m=5, method = 'pmm', seed = 101)

# miceDoctorate <- complete(imputed\_Doctorate,3)

# ProtectedLand$Doctorate <- miceDoctorate$ProtectedLand.Doctorate

#

# any(is.na(ProtectedLand$Doctorate))

#

# #Clearing unneeded object

# rm(imputed\_Doctorate)

# rm(miceDoctorate)

#

# ######Using R Mice Package to impute missing values [Voted]

# any(is.na(ProtectedLand$Voted))

#

# miceVoted <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$Voted)

# md.pattern(miceVoted)

# imputed\_Voted <- mice(miceVoted, m=5, method = 'pmm', seed = 101)

# miceVoted <- complete(imputed\_Voted,3)

# ProtectedLand$Voted <- miceVoted$ProtectedLand.Voted

#

# any(is.na(ProtectedLand$Voted))

#

# #Clearing unneeded object

# rm(imputed\_Voted)

# rm(miceVoted)

#

# ######Using R Mice Package to impute missing values [Registered]

# any(is.na(ProtectedLand$Registered))

#

# miceRegistered <- data.frame(ProtectedLand$census\_tract\_population, ProtectedLand$Registered)

# md.pattern(miceRegistered)

# imputed\_Registered <- mice(miceRegistered, m=5, method = 'pmm', seed = 101)

# miceRegistered <- complete(imputed\_Registered,3)

# ProtectedLand$Registered <- miceRegistered$ProtectedLand.Registered

#

# any(is.na(ProtectedLand$Registered))

#

# #Clearing unneeded object

# rm(imputed\_Registered)

# rm(miceRegistered)

#

# # Clear the "Values" section

# rm(list=ls.str(mode='numeric'))

#--------------------------------------Descriptive Analysis [i003]--------------------------------------------------------------

#Generate a vector of unique counties

uniqueCounty <- unique(ProtectedLand$county)

# Create new fields calculating the population of the tract id based on ethnicity

ProtectedLand$number\_white <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$share\_white,0)

ProtectedLand$number\_black <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$share\_black,0)

ProtectedLand$number\_native\_american <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$share\_native\_american,0)

ProtectedLand$number\_asian <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$share\_asian,0)

ProtectedLand$number\_pacific\_islander <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$share\_pacific\_islander,0)

ProtectedLand$number\_other <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$share\_other,0)

ProtectedLand$number\_2plus <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$share\_2plus,0)

ProtectedLand$number\_hispanic <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$share\_hispanic,0)

# Create new fields calculating the population of the tract id based on education

ProtectedLand$number\_lessthanHS <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$lessthanHS,0)

ProtectedLand$number\_hsDiploma <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$HSdiploma,0)

ProtectedLand$number\_someCollege <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$someCollege,0)

ProtectedLand$number\_Associates <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$Associates,0)

ProtectedLand$number\_Bachelors <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$Bachelors,0)

ProtectedLand$number\_Masters <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$Masters,0)

ProtectedLand$number\_Professional <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$Professional,0)

ProtectedLand$number\_Doctorate <- round(ProtectedLand$census\_tract\_population \* ProtectedLand$Doctorate,0)

# Create new fields calculating the population of the tract id based on education

ProtectedLand$number\_Registered <- round(ProtectedLand$Registered \* ProtectedLand$Registered....of.Eligible.,0)

ProtectedLand$number\_Democrat <- round(ProtectedLand$Registered \* ProtectedLand$Democratic....of.Eligible.,0)

ProtectedLand$number\_Republican <- round(ProtectedLand$Registered \* ProtectedLand$Republican....of.Eligible.,0)

ProtectedLand$number\_Independent <- round(ProtectedLand$Registered \* ProtectedLand$American.Independent....of.Eligible.,0)

ProtectedLand$number\_Green <- round(ProtectedLand$Registered \* ProtectedLand$Green....of.Eligible.,0)

ProtectedLand$number\_Libertarian <- round(ProtectedLand$Registered \* ProtectedLand$Libertarian....of.Eligible.,0)

ProtectedLand$number\_PeaceAndFreedom <- round(ProtectedLand$Registered \* ProtectedLand$Peace.and.Freedom....of.Eligible.,0)

ProtectedLand$number\_Party\_Other <- round(ProtectedLand$Registered \* ProtectedLand$Other....of.Eligible.,0)

ProtectedLand$number\_Party\_Declined <- round(ProtectedLand$Registered \* ProtectedLand$Decline.to.State....of.Eligible.,0)

# Create a function that generates a new data frame that houses all descriptive stats

myDescriptiveStats <- data.frame()

getDescriptiveStats <- function(){

for ( iter in 1:length(uniqueCounty)) {

myCounty <- uniqueCounty[iter]

mean\_ct\_pop <- round(mean(ProtectedLand$census\_tract\_population[ProtectedLand$county==myCounty]),0)

median\_ct\_pop <- round(median(ProtectedLand$census\_tract\_population[ProtectedLand$county==myCounty]),0)

min\_ct\_pop <- round(min(ProtectedLand$census\_tract\_population[ProtectedLand$county==myCounty]),0)

max\_ct\_pop <- round(max(ProtectedLand$census\_tract\_population[ProtectedLand$county==myCounty]),0)

mean\_income <- round(mean(ProtectedLand$average\_income[ProtectedLand$county==myCounty]),0)

median\_income <- round(median(ProtectedLand$average\_income[ProtectedLand$county==myCounty]),0)

min\_income <- round(min(ProtectedLand$average\_income[ProtectedLand$county==myCounty]),0)

max\_income <- round(max(ProtectedLand$average\_income[ProtectedLand$county==myCounty]),0)

mean\_housing\_price <- round(mean(ProtectedLand$median\_housing\_price[ProtectedLand$county==myCounty]),0)

median\_housing\_price <- round(median(ProtectedLand$median\_housing\_price[ProtectedLand$county==myCounty]),0)

min\_housing\_price <- round(min(ProtectedLand$median\_housing\_price[ProtectedLand$county==myCounty]),0)

max\_housing\_price <- round(max(ProtectedLand$median\_housing\_price[ProtectedLand$county==myCounty]),0)

mean\_white\_pop <- round(mean(ProtectedLand$number\_white[ProtectedLand$county==myCounty]),0)

median\_white\_pop <- round(median(ProtectedLand$number\_white[ProtectedLand$county==myCounty]),0)

min\_white\_pop <- round(min(ProtectedLand$number\_white[ProtectedLand$county==myCounty]),0)

max\_white\_pop <- round(max(ProtectedLand$number\_white[ProtectedLand$county==myCounty]),0)

mean\_black\_pop <- round(mean(ProtectedLand$number\_black[ProtectedLand$county==myCounty]),0)

median\_black\_pop <- round(median(ProtectedLand$number\_black[ProtectedLand$county==myCounty]),0)

min\_black\_pop <- round(min(ProtectedLand$number\_black[ProtectedLand$county==myCounty]),0)

max\_black\_pop <- round(max(ProtectedLand$number\_black[ProtectedLand$county==myCounty]),0)

mean\_native\_american\_pop <- round(mean(ProtectedLand$number\_native\_american[ProtectedLand$county==myCounty]),0)

median\_native\_american\_pop <- round(median(ProtectedLand$number\_native\_american[ProtectedLand$county==myCounty]),0)

min\_native\_american\_pop <- round(min(ProtectedLand$number\_native\_american[ProtectedLand$county==myCounty]),0)

max\_native\_american\_pop <- round(max(ProtectedLand$number\_native\_american[ProtectedLand$county==myCounty]),0)

mean\_asian\_pop <- round(mean(ProtectedLand$number\_asian[ProtectedLand$county==myCounty]),0)

median\_asian\_pop <- round(median(ProtectedLand$number\_asian[ProtectedLand$county==myCounty]),0)

min\_asian\_pop <- round(min(ProtectedLand$number\_asian[ProtectedLand$county==myCounty]),0)

max\_asian\_pop <- round(max(ProtectedLand$number\_asian[ProtectedLand$county==myCounty]),0)

mean\_pacific\_islander\_pop <- round(mean(ProtectedLand$number\_pacific\_islander[ProtectedLand$county==myCounty]),0)

median\_pacific\_islander\_pop <- round(median(ProtectedLand$number\_pacific\_islander[ProtectedLand$county==myCounty]),0)

min\_pacific\_islander\_pop <- round(min(ProtectedLand$number\_pacific\_islander[ProtectedLand$county==myCounty]),0)

max\_pacific\_islander\_pop <- round(max(ProtectedLand$number\_pacific\_islander[ProtectedLand$county==myCounty]),0)

mean\_other\_pop <- round(mean(ProtectedLand$number\_other[ProtectedLand$county==myCounty]),0)

median\_other\_pop <- round(median(ProtectedLand$number\_other[ProtectedLand$county==myCounty]),0)

min\_other\_pop <- round(min(ProtectedLand$number\_other[ProtectedLand$county==myCounty]),0)

max\_other\_pop <- round(max(ProtectedLand$number\_other[ProtectedLand$county==myCounty]),0)

mean\_2plus\_pop <- round(mean(ProtectedLand$number\_2plus[ProtectedLand$county==myCounty]),0)

median\_2plus\_pop <- round(median(ProtectedLand$number\_2plus[ProtectedLand$county==myCounty]),0)

min\_2plus\_pop <- round(min(ProtectedLand$number\_2plus[ProtectedLand$county==myCounty]),0)

max\_2plus\_pop <- round(max(ProtectedLand$number\_2plus[ProtectedLand$county==myCounty]),0)

mean\_lessthanHS <- round(mean(ProtectedLand$number\_lessthanHS[ProtectedLand$county==myCounty]),0)

median\_lessthanHS <- round(mean(ProtectedLand$number\_lessthanHS[ProtectedLand$county==myCounty]),0)

min\_lessthanHS <- round(min(ProtectedLand$number\_lessthanHS[ProtectedLand$county==myCounty]),0)

max\_lessthanHS <- round(max(ProtectedLand$number\_lessthanHS[ProtectedLand$county==myCounty]),0)

mean\_hsDiploma <- round(mean(ProtectedLand$number\_hsDiploma[ProtectedLand$county==myCounty]),0)

median\_hsDiploma <- round(mean(ProtectedLand$number\_hsDiploma[ProtectedLand$county==myCounty]),0)

min\_hsDiploma <- round(min(ProtectedLand$number\_hsDiploma[ProtectedLand$county==myCounty]),0)

max\_hsDiploma <- round(max(ProtectedLand$number\_hsDiploma[ProtectedLand$county==myCounty]),0)

mean\_someCollege <- round(mean(ProtectedLand$number\_someCollege[ProtectedLand$county==myCounty]),0)

median\_someCollege <- round(mean(ProtectedLand$number\_someCollege[ProtectedLand$county==myCounty]),0)

min\_someCollege <- round(min(ProtectedLand$number\_someCollege[ProtectedLand$county==myCounty]),0)

max\_someCollege <- round(max(ProtectedLand$number\_someCollege[ProtectedLand$county==myCounty]),0)

mean\_Associates <- round(mean(ProtectedLand$number\_Associates[ProtectedLand$county==myCounty]),0)

median\_Associates <- round(mean(ProtectedLand$number\_Associates[ProtectedLand$county==myCounty]),0)

min\_Associates <- round(min(ProtectedLand$number\_Associates[ProtectedLand$county==myCounty]),0)

max\_Associates <- round(max(ProtectedLand$number\_Associates[ProtectedLand$county==myCounty]),0)

mean\_Bachelors <- round(mean(ProtectedLand$number\_Bachelors[ProtectedLand$county==myCounty]),0)

median\_Bachelors <- round(mean(ProtectedLand$number\_Bachelors[ProtectedLand$county==myCounty]),0)

min\_Bachelors <- round(min(ProtectedLand$number\_Bachelors[ProtectedLand$county==myCounty]),0)

max\_Bachelors <- round(max(ProtectedLand$number\_Bachelors[ProtectedLand$county==myCounty]),0)

mean\_Masters <- round(mean(ProtectedLand$number\_Masters[ProtectedLand$county==myCounty]),0)

median\_Masters <- round(mean(ProtectedLand$number\_Masters[ProtectedLand$county==myCounty]),0)

min\_Masters <- round(min(ProtectedLand$number\_Masters[ProtectedLand$county==myCounty]),0)

max\_Masters <- round(max(ProtectedLand$number\_Masters[ProtectedLand$county==myCounty]),0)

mean\_Professional <- round(mean(ProtectedLand$number\_Professional[ProtectedLand$county==myCounty]),0)

median\_Professional <- round(mean(ProtectedLand$number\_Professional[ProtectedLand$county==myCounty]),0)

min\_Professional <- round(min(ProtectedLand$number\_Professional[ProtectedLand$county==myCounty]),0)

max\_Professional <- round(max(ProtectedLand$number\_Professional[ProtectedLand$county==myCounty]),0)

mean\_Doctorate <- round(mean(ProtectedLand$number\_Doctorate[ProtectedLand$county==myCounty]),0)

median\_Doctorate <- round(mean(ProtectedLand$number\_Doctorate[ProtectedLand$county==myCounty]),0)

min\_Doctorate <- round(min(ProtectedLand$number\_Doctorate[ProtectedLand$county==myCounty]),0)

max\_Doctorate <- round(max(ProtectedLand$number\_Doctorate[ProtectedLand$county==myCounty]),0)

mean\_Voted <- round(mean(ProtectedLand$Voted[ProtectedLand$county==myCounty]),0)

median\_Voted <- round(mean(ProtectedLand$Voted[ProtectedLand$county==myCounty]),0)

min\_Voted <- round(min(ProtectedLand$Voted[ProtectedLand$county==myCounty]),0)

max\_Voted <- round(max(ProtectedLand$Voted[ProtectedLand$county==myCounty]),0)

mean\_Registered <- round(mean(ProtectedLand$Registered[ProtectedLand$county==myCounty]),0)

median\_Registered <- round(mean(ProtectedLand$Registered[ProtectedLand$county==myCounty]),0)

min\_Registered <- round(min(ProtectedLand$Registered[ProtectedLand$county==myCounty]),0)

max\_Registered <- round(max(ProtectedLand$Registered[ProtectedLand$county==myCounty]),0)

mean\_voting\_age\_pop <- round(mean(ProtectedLand$voting\_age\_pop[ProtectedLand$county==myCounty]),0)

median\_voting\_age\_pop <- round(mean(ProtectedLand$voting\_age\_pop[ProtectedLand$county==myCounty]),0)

min\_voting\_age\_pop <- round(min(ProtectedLand$voting\_age\_pop[ProtectedLand$county==myCounty]),0)

max\_voting\_age\_pop <- round(max(ProtectedLand$voting\_age\_pop[ProtectedLand$county==myCounty]),0)

mean\_number\_democrat <- round(mean(ProtectedLand$number\_Democrat[ProtectedLand$county==myCounty]),0)

median\_number\_democrat <- round(mean(ProtectedLand$number\_Democrat[ProtectedLand$county==myCounty]),0)

min\_number\_democrat <- round(min(ProtectedLand$number\_Democrat[ProtectedLand$county==myCounty]),0)

max\_number\_democrat <- round(max(ProtectedLand$number\_Democrat[ProtectedLand$county==myCounty]),0)

mean\_number\_republican <- round(mean(ProtectedLand$number\_Republican[ProtectedLand$county==myCounty]),0)

median\_number\_republican <- round(mean(ProtectedLand$number\_Republican[ProtectedLand$county==myCounty]),0)

min\_number\_republican <- round(min(ProtectedLand$number\_Republican[ProtectedLand$county==myCounty]),0)

max\_number\_republican <- round(max(ProtectedLand$number\_Republican[ProtectedLand$county==myCounty]),0)

mean\_number\_independent <- round(mean(ProtectedLand$number\_Independent[ProtectedLand$county==myCounty]),0)

median\_number\_independent <- round(mean(ProtectedLand$number\_Independent[ProtectedLand$county==myCounty]),0)

min\_number\_independent <- round(min(ProtectedLand$number\_Independent[ProtectedLand$county==myCounty]),0)

max\_number\_independent <- round(max(ProtectedLand$number\_Independent[ProtectedLand$county==myCounty]),0)

mean\_number\_green <- round(mean(ProtectedLand$number\_Green[ProtectedLand$county==myCounty]),0)

median\_number\_green <- round(mean(ProtectedLand$number\_Green[ProtectedLand$county==myCounty]),0)

min\_number\_green <- round(min(ProtectedLand$number\_Green[ProtectedLand$county==myCounty]),0)

max\_number\_green <- round(max(ProtectedLand$number\_Green[ProtectedLand$county==myCounty]),0)

mean\_number\_libertarian <- round(mean(ProtectedLand$number\_Libertarian[ProtectedLand$county==myCounty]),0)

median\_number\_libertarian <- round(mean(ProtectedLand$number\_Libertarian[ProtectedLand$county==myCounty]),0)

min\_number\_libertarian <- round(min(ProtectedLand$number\_Libertarian[ProtectedLand$county==myCounty]),0)

max\_number\_libertarian <- round(max(ProtectedLand$number\_Libertarian[ProtectedLand$county==myCounty]),0)

mean\_number\_peaceandfreedom <- round(mean(ProtectedLand$number\_PeaceAndFreedom[ProtectedLand$county==myCounty]),0)

median\_number\_peaceandfreedom <- round(mean(ProtectedLand$number\_PeaceAndFreedom[ProtectedLand$county==myCounty]),0)

min\_number\_peaceandfreedom <- round(min(ProtectedLand$number\_PeaceAndFreedom[ProtectedLand$county==myCounty]),0)

max\_number\_peaceandfreedom <- round(max(ProtectedLand$number\_PeaceAndFreedom[ProtectedLand$county==myCounty]),0)

mean\_number\_party\_other <- round(mean(ProtectedLand$number\_Party\_Other[ProtectedLand$county==myCounty]),0)

median\_number\_party\_other <- round(mean(ProtectedLand$number\_Party\_Other[ProtectedLand$county==myCounty]),0)

min\_number\_party\_other <- round(min(ProtectedLand$number\_Party\_Other[ProtectedLand$county==myCounty]),0)

max\_number\_party\_other <- round(max(ProtectedLand$number\_Party\_Other[ProtectedLand$county==myCounty]),0)

mean\_number\_party\_declined <- round(mean(ProtectedLand$number\_Party\_Declined[ProtectedLand$county==myCounty]),0)

median\_number\_party\_declined <- round(mean(ProtectedLand$number\_Party\_Declined[ProtectedLand$county==myCounty]),0)

min\_number\_party\_declined <- round(min(ProtectedLand$number\_Party\_Declined[ProtectedLand$county==myCounty]),0)

max\_number\_party\_declined <- round(max(ProtectedLand$number\_Party\_Declined[ProtectedLand$county==myCounty]),0)

newRow <- data.frame(county=myCounty,

mean\_ct\_pop, median\_ct\_pop, min\_ct\_pop, max\_ct\_pop,

mean\_income, median\_income, min\_income, max\_income,

mean\_housing\_price, median\_housing\_price, min\_housing\_price, max\_housing\_price,

mean\_white\_pop,median\_white\_pop,min\_white\_pop,max\_white\_pop,

mean\_black\_pop,median\_black\_pop,min\_black\_pop,max\_black\_pop,

mean\_native\_american\_pop,median\_native\_american\_pop,min\_native\_american\_pop,max\_native\_american\_pop,

mean\_asian\_pop,median\_asian\_pop,min\_asian\_pop,max\_asian\_pop,

mean\_pacific\_islander\_pop,median\_pacific\_islander\_pop,min\_pacific\_islander\_pop,max\_pacific\_islander\_pop,

mean\_other\_pop,median\_other\_pop,min\_other\_pop,max\_other\_pop,

mean\_2plus\_pop,median\_2plus\_pop,min\_2plus\_pop,max\_2plus\_pop,

mean\_lessthanHS, median\_lessthanHS, min\_lessthanHS, max\_lessthanHS,

mean\_hsDiploma, median\_hsDiploma, min\_hsDiploma, max\_hsDiploma,

mean\_someCollege, median\_someCollege, min\_someCollege, max\_someCollege,

mean\_Associates, median\_Associates, min\_Associates, max\_Associates,

mean\_Bachelors, median\_Bachelors, min\_Bachelors, max\_Bachelors,

mean\_Masters, median\_Masters, min\_Masters, max\_Masters,

mean\_Professional, median\_Professional, min\_Professional, max\_Professional,

mean\_Doctorate, median\_Doctorate, min\_Doctorate, max\_Doctorate,

mean\_Voted, median\_Voted, min\_Voted, max\_Voted,

mean\_Registered, median\_Registered, min\_Registered, max\_Registered,

mean\_voting\_age\_pop, median\_voting\_age\_pop, min\_voting\_age\_pop, max\_voting\_age\_pop,

mean\_number\_democrat, median\_number\_democrat, min\_number\_democrat, max\_number\_democrat,

mean\_number\_republican, median\_number\_republican, min\_number\_republican, max\_number\_republican,

mean\_number\_independent, median\_number\_independent, min\_number\_independent, max\_number\_independent,

mean\_number\_green, median\_number\_green, min\_number\_green, max\_number\_green,

mean\_number\_libertarian, median\_number\_libertarian, min\_number\_libertarian, max\_number\_libertarian,

mean\_number\_peaceandfreedom, median\_number\_peaceandfreedom, min\_number\_peaceandfreedom, max\_number\_peaceandfreedom,

mean\_number\_party\_other, median\_number\_party\_other, min\_number\_party\_other, max\_number\_party\_other,

mean\_number\_party\_declined, median\_number\_party\_declined, min\_number\_party\_declined, max\_number\_party\_declined

)

myDescriptiveStats <<- rbind(myDescriptiveStats,newRow)

}

}

getDescriptiveStats()

#----------------------------------------Descriptive Charts [i003a]----------------------------------------

#Generate lolipop chart to show population distribution by county

# Create a subset of data and save it to a dataframe

myPopTotalsByCounty <- ProtectedLand[,c("county","county\_population")]

# Deduplicate the rows

myPopTotalsByCounty <- myPopTotalsByCounty[!duplicated(myPopTotalsByCounty[, c("county","county\_population")]), ]

# Reset the row names

row.names(myPopTotalsByCounty) <- NULL

# Plot the data

theme\_set(theme\_bw())

ggplot(myPopTotalsByCounty, aes(x=county, y=county\_population)) +

geom\_point(size=3) +

geom\_segment(aes(x=county, xend=county, y=0, yend=county\_population)) +

labs(title="Population vs County", caption="LA county accounts for ~27% of Californias population.") +

theme(axis.text.x=element\_text(angle=65, hjust=1, vjust=1))

#Generate box charts to visualize ethnic and education dispersion in California

# Create a subset of data and save it to a dataframe

myEthnicData <- ProtectedLand[,c("number\_white","number\_black",

"number\_native\_american","number\_asian",

"number\_pacific\_islander","number\_other",

"number\_2plus","number\_hispanic")]

myEducationData <- ProtectedLand[,c("number\_lessthanHS","number\_hsDiploma",

"number\_someCollege","number\_Associates",

"number\_Bachelors","number\_Masters",

"number\_Professional","number\_Doctorate")]

# Melt the data for easier plotting

myMeltedEthnicData <- melt(data=myEthnicData,measure.vars = c("number\_white","number\_black",

"number\_native\_american","number\_asian",

"number\_pacific\_islander","number\_other",

"number\_2plus","number\_hispanic"))

myMeltedEducationData <- melt(data=myEducationData,measure.vars = c("number\_lessthanHS","number\_hsDiploma",

"number\_someCollege","number\_Associates",

"number\_Bachelors","number\_Masters",

"number\_Professional","number\_Doctorate"))

# Relabel column headers and remove "number\_"

names(myMeltedEthnicData)[1] <- "Ethnicity"

names(myMeltedEthnicData)[2] <- "Population"

myMeltedEthnicData$Ethnicity <- gsub("number\_", "", myMeltedEthnicData$Ethnicity)

names(myMeltedEducationData)[1] <- "Education"

names(myMeltedEducationData)[2] <- "Population"

myMeltedEducationData$Education <- gsub("number\_", "", myMeltedEducationData$Education)

# Make additional column so plot order mimics dataframe column order

myMeltedEthnicData$Ethnicity2 <- factor(myMeltedEthnicData$Ethnicity, c("white","black",

"native\_american","asian",

"pacific\_islander","other",

"2plus","hispanic"))

myMeltedEducationData$Education2 <- factor(myMeltedEducationData$Education, c("lessthanHS","hsDiploma",

"someCollege","Associates",

"Bachelors","Masters",

"Professional","Doctorate"))

# Box-plot ethnicity, education and income in california

ethnicBoxData <- ggplot(data=myMeltedEthnicData, aes(x=Ethnicity2,y=Population,color=Ethnicity))

ethnicBoxData + geom\_boxplot(size=1) + theme(legend.position = "none", plot.title=element\_text(hjust=0.5)) +

labs(x="race",y="population") + ggtitle("population vs race")

print("White and hispanic population distributions are very similar with white groups having a slightly

higher median with one notable outlier enclave (Los Angeles).")

educationBoxData <- ggplot(data=myMeltedEducationData, aes(x=Education2,y=Population,color=Education))

educationBoxData + geom\_boxplot(size=1) + theme(legend.position = "none", plot.title=element\_text(hjust=0.5)) +

labs(x="education",y="population") + ggtitle("population vs education level")

print("Educational dispersion is very similar between high school diploma and some college, with Bachelors

and less than high school showing greater dispersion and less consistancy around the median.")

meanAverageIncome <- data.frame(round(mean(ProtectedLand$average\_income, na.rm = T),0))

names(meanAverageIncome)[1] <- "average\_income"

incomeBoxData <- ggplot(data=ProtectedLand, aes(x="",y=average\_income))

incomeBoxData + geom\_boxplot(size=1) +

stat\_summary(fun.y = mean, geom = "point", shape=23, size=4) +

geom\_text(data = meanAverageIncome, aes(label = average\_income), nudge\_x = .03, vjust=-0.5) +

theme(legend.position = "none", plot.title=element\_text(hjust=0.5)) +

labs(x="",y="average income") + ggtitle("average income dispersion in california")

print("Most people in California make between $20,000 and $40,000 on average. There is a considerable spread of average salaries.")

#----------------------------------------Descriptive Maps [i003b]----------------------------------------

mappingData<-data.frame(mappingData)

###Registered to eligible voters [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$registered\_to\_eligible)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$registered\_to\_eligible))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Registered to Eligible Voters (Dist2Tract)",fill="% Registered to Eligible Voters")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Registered to eligible voters [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$registered\_to\_eligible)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$registered\_to\_eligible))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+labs(title = "Registered to Eligible Voters (ldist)", fill="% Registered to Eligible Voters")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Democratic to registered voters [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$democratic\_to\_registered)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$democratic\_to\_registered))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Democratic to Registered Voters (Dist2Tract)",fill="% Democratic to Registered Voters")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Democratic to registered voters [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$democratic\_to\_registered)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$democratic\_to\_registered))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+labs(title = "Democratic to Registered Voters (ldist)", fill="% Democratic to Registered Voters")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Republican to registered voters [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$republican\_to\_registered)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$republican\_to\_registered))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Republican to Registered Voters (Dist2Tract)",fill="% Republican to Registered Voters")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Republican to registered voters [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$republican\_to\_registered)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$republican\_to\_registered))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Republican to Registered Voters (ldist)", fill="% Republican to Registered Voters")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###No Party to registered voters [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$noparty\_to\_registered)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$noparty\_to\_registered))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "No Party to Registered Voters (Dist2Tract)",fill="% No Party to Registered Voters")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###No Party to registered voters [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$noparty\_to\_registered)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$noparty\_to\_registered))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "No Party to Registered Voters (ldist)", fill="% No Party to Registered Voters")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Average Income [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$avgincome)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$avgincome))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Average Income (Dist2Tract)",fill="Average Income")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Average Income [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$avgincome)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$avgincome))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Average Income (ldist)", fill="Average Income")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Median Housing Price [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$med\_housing)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$med\_housing))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Median Housing Price (Dist2Tract)",fill="Median Housing Price")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Median Housing Price[county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$med\_housing)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$med\_housing))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Median Housing Price (ldist)", fill="Median Housing Price")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###County Population [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$Cpop)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$Cpop))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "County Population (Dist2Tract)",fill="Median Housing Price")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###County Population [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$Cpop)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$Cpop))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "County Population (ldist)", fill="County Population")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Population Density [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$popdens)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$popdens))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Population Density (Dist2Tract)",fill="Population Density")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Population Density [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$popdens)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$popdens))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Population Density (ldist)", fill="Population Density")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Share White [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$share\_white)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$share\_white))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Share White (Dist2Tract)",fill="Share White")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Share White [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$share\_white)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$share\_white))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Share White (ldist)", fill="Share White")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Share Black [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$share\_black)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$share\_black))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Share Black (Dist2Tract)",fill="Share Black")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Share Black [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$share\_black)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$share\_black))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Share Black (ldist)", fill="Share Black")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Share Hispanic [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$share\_hispanic)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$share\_hispanic))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Share Hispanic (Dist2Tract)",fill="Share Hispanic")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Share Hispanic [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$share\_hispanic)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$share\_hispanic))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Share Hispanic (ldist)", fill="Share Hispanic")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Share Asian [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$share\_asian)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$share\_asian))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Share Asian (Dist2Tract)",fill="Share Asian")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

#Share Asian[county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$share\_asian)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$share\_asian))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Share Asian (ldist)", fill="Share Asian")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Less than HS Education [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$lessthanHS)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$lessthanHS))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Less than HS Education (Dist2Tract)",fill="Less than HS Education")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

#Less than HS Education [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$lessthanHS)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$lessthanHS))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Less than HS Education (ldist)", fill="Less than HS Education")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Professional Degrees [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$Professional)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$Professional))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Professional Degrees (Dist2Tract)",fill="Professional Degrees")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

#Professional Degrees [county\_ldist]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$Professional)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$Professional))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Professional Degrees (ldist)", fill="Professional Degrees")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Doctorate [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$Doctorate)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$Doctorate))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_dist2tract, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_dist2tract", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Doctorate (Dist2Tract)",fill="Doctorate")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

###Doctorate [county\_dist2tract]

ww <- ggplot(mappingData, aes(x=long, y=lat, group=group, fill = mappingData$Doctorate)) + geom\_polygon(colour="black")+ coord\_map('polyconic')

xx <- ww+scale\_fill\_gradient2(low="#559999", mid="grey90", high="#BB650B", midpoint=median(mappingData$Doctorate))

yy <- xx +geom\_point( data=mappingData, aes(x=long, y=lat, size = mappingData$county\_ldist, color="hotpink2", alpha = 0.1)) + scale\_size\_continuous(name="mappingData$county\_ldist", range = c(1,10)) + guides(colour=FALSE, alpha=FALSE) #dots

zz <- yy+scale\_fill\_viridis\_c()+labs(title = "Doctorate (ldist)",fill="Doctorate")

zz + theme(axis.text.x = element\_blank(), axis.text.y = element\_blank(), axis.ticks = element\_blank(),rect = element\_blank(), axis.title = element\_blank(), panel.background = element\_blank(), panel.grid.major = element\_blank(), plot.title=element\_text(hjust=0.5))

#--------------------------------------Advanced Analysis [i004]--------------------------------------------------------------

hist(ProtectedLand$distance\_to\_tract)

hist(ProtectedLand$ldist)

print("Using the log of distance instead of the regular distance variable normalizes the data, allowing for more accurate linear models.")

# 1. Can we show inequality in access to protected land areas?

# 1a. Is there a difference in access for people in rural versus suburban versus urban counties?

model1a <- lm(ldist ~ urban + suburb, ProtectedLand)

summary(model1a)

print("1a. Urban and suburban counties have increased access to protected land areas compared to rural areas. The model has a near-0 p-value and an adjusted R-squared of 11.21%. Suburban county census tracts see a exp(-1.13181)-1 = 67.75% increase in access and urban county census tracts see a 78.48% increase in access.")

model1a\_urban <- ggplot(model1a$model, aes\_string(x = names(model1a$model)[2], y = names(model1a$model)[1])) + geom\_point(colour="lightblue", alpha = 0.01) + geom\_abline(intercept = coef(model1a)[1], slope = coef(model1a)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1a\_suburb <- ggplot(model1a$model, aes\_string(x = names(model1a$model)[3], y = names(model1a$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.01) + geom\_abline(intercept = coef(model1a)[1], slope = coef(model1a)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1a\_urban, model1a\_suburb, nrow=2, top="Increased access for urban and suburban areas compared to rural areas")

# 1ai. Is there a relationship between rural, suburban, and urban incomes?

model1ai <- lm(average\_income ~ urban + suburb, ProtectedLand)

summary(model1ai)

print("1ai. The model is significant and has an adjusted R-squared value of 2.784%. Both coefficients are significant, showing an increase of approximately $10,000 for urban and $4,500 for suburban in average income over rural incomes.")

model1ai\_urban <- ggplot(model1ai$model, aes\_string(x = names(model1ai$model)[2], y = names(model1ai$model)[1])) + geom\_point(colour="lightblue", alpha = 0.01) + geom\_abline(intercept = coef(model1ai)[1], slope = coef(model1ai)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1ai\_suburb <- ggplot(model1ai$model, aes\_string(x = names(model1ai$model)[3], y = names(model1ai$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.01) + geom\_abline(intercept = coef(model1ai)[1], slope = coef(model1ai)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1ai\_urban, model1ai\_suburb, nrow=2, top="Increased income for urban and suburban areas compared to rural areas")

# 1aia. Does that factor into median housing price?

model1aia <- lm(median\_housing\_price ~ average\_income + urban + suburb, ProtectedLand)

summary(model1aia)

print("1aia. The model is significant and has an adjusted R-squared of 66.23%. Average income and urban locations seem to be the primary drivers of housing prices in California, although suburban housing prices are also much higher.")

model1aia\_avgInc <- ggplot(model1aia$model, aes\_string(x = names(model1aia$model)[2], y = names(model1aia$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model1aia)[1], slope = coef(model1aia)[2], colour="black", size=1)

model1aia\_urban <- ggplot(model1aia$model, aes\_string(x = names(model1aia$model)[3], y = names(model1aia$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.01) + geom\_abline(intercept = coef(model1aia)[1], slope = coef(model1aia)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1aia\_suburb <- ggplot(model1aia$model, aes\_string(x = names(model1aia$model)[4], y = names(model1aia$model)[1])) + geom\_point(colour="firebrick", alpha = 0.01) + geom\_abline(intercept = coef(model1aia)[1], slope = coef(model1aia)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1aia\_avgInc, model1aia\_urban, model1aia\_suburb, nrow=3, top="Increased median housing price for urban and suburban areas compared to rural areas")

# 1b. Is there a relationship between race and access?

model1b <- lm(ldist ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLand)

summary(model1b)

print("1b. The model is significant and has an adjusted R-squared of 9.437%. Most race/ethnicity coefficients are statistically significant and show slightly increased access compared to white populations. Of notable exception is that Native American populations see highly decreased access (increased distance), which is expected since Native land is not included in the group of designated protected lands.")

model1b\_black <- ggplot(model1b$model, aes\_string(x = names(model1b$model)[2], y = names(model1b$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model1b)[1], slope = coef(model1b)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1b\_hispanic <- ggplot(model1b$model, aes\_string(x = names(model1b$model)[3], y = names(model1b$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1b)[1], slope = coef(model1b)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1b\_asian <- ggplot(model1b$model, aes\_string(x = names(model1b$model)[4], y = names(model1b$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1b)[1], slope = coef(model1b)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1b\_na <- ggplot(model1b$model, aes\_string(x = names(model1b$model)[5], y = names(model1b$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1b)[1], slope = coef(model1b)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1b\_pi <- ggplot(model1b$model, aes\_string(x = names(model1b$model)[6], y = names(model1b$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1b)[1], slope = coef(model1b)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1b\_2plus <- ggplot(model1b$model, aes\_string(x = names(model1b$model)[7], y = names(model1b$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model1b)[1], slope = coef(model1b)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1b\_other <- ggplot(model1b$model, aes\_string(x = names(model1b$model)[8], y = names(model1b$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model1b)[1], slope = coef(model1b)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1b\_black, model1b\_hispanic, model1b\_asian, model1b\_na, model1b\_pi, model1b\_2plus, model1b\_other, nrow=4, top="Most racial groups show slightly increased access to protected lands compared to white people")

# 1bi. Is there a relationship between race and income?

model1bi <- lm(average\_income ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLand)

summary(model1bi)

print("1bi. The model is significant and has an adjusted R-squared of 54.05%. All minority racial/ethnic populations have lower average incomes compared to white populations except those that identify as other. The coefficients can be interpreted as a 1 percentage point (0.01) increase in the population share of a racial/ethnic minority in a census tract will show a $[coefficient value] change in census tract average income.")

model1bi\_black <- ggplot(model1bi$model, aes\_string(x = names(model1bi$model)[2], y = names(model1bi$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model1bi)[1], slope = coef(model1bi)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1bi\_hispanic <- ggplot(model1bi$model, aes\_string(x = names(model1bi$model)[3], y = names(model1bi$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1bi)[1], slope = coef(model1bi)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1bi\_asian <- ggplot(model1bi$model, aes\_string(x = names(model1bi$model)[4], y = names(model1bi$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1bi)[1], slope = coef(model1bi)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1bi\_na <- ggplot(model1bi$model, aes\_string(x = names(model1bi$model)[5], y = names(model1bi$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1bi)[1], slope = coef(model1bi)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1bi\_pi <- ggplot(model1bi$model, aes\_string(x = names(model1bi$model)[6], y = names(model1bi$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1bi)[1], slope = coef(model1bi)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1bi\_2plus <- ggplot(model1bi$model, aes\_string(x = names(model1bi$model)[7], y = names(model1bi$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model1bi)[1], slope = coef(model1bi)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1bi\_other <- ggplot(model1bi$model, aes\_string(x = names(model1bi$model)[8], y = names(model1bi$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model1bi)[1], slope = coef(model1bi)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

grid.arrange(model1bi\_black,model1bi\_hispanic,model1bi\_asian,model1bi\_na,model1bi\_pi,model1bi\_2plus,model1bi\_other, nrow=4, top="All minority groups have lower average incomes compared to their white counterparts")

# 1bii. Is there a relationship between race and housing price?

model1bii <- lm(median\_housing\_price ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLand)

summary(model1bii)

print("1bii. The model is significant and has an adjusted R-squared of 39.59%. All minority racial/ethnic populations have lower median housing prices compared to white populations except those that identify as Asian or other. The coefficients can be interpreted as a 1 percentage point (0.01) increase in the population share of a racial/ethnic minority in a census tract will show a $[coefficient value] change in census tract median housing price.")

model1bii\_black <- ggplot(model1bii$model, aes\_string(x = names(model1bii$model)[2], y = names(model1bii$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model1bii)[1], slope = coef(model1bii)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1bii\_hispanic <- ggplot(model1bii$model, aes\_string(x = names(model1bii$model)[3], y = names(model1bii$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1bii)[1], slope = coef(model1bii)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1bii\_asian <- ggplot(model1bii$model, aes\_string(x = names(model1bii$model)[4], y = names(model1bii$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1bii)[1], slope = coef(model1bii)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1bii\_na <- ggplot(model1bii$model, aes\_string(x = names(model1bii$model)[5], y = names(model1bii$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1bii)[1], slope = coef(model1bii)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1bii\_pi <- ggplot(model1bii$model, aes\_string(x = names(model1bii$model)[6], y = names(model1bii$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1bii)[1], slope = coef(model1bii)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1bii\_2plus <- ggplot(model1bii$model, aes\_string(x = names(model1bii$model)[7], y = names(model1bii$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model1bii)[1], slope = coef(model1bii)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1bii\_other <- ggplot(model1bii$model, aes\_string(x = names(model1bii$model)[8], y = names(model1bii$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model1bii)[1], slope = coef(model1bii)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

grid.arrange(model1bii\_black,model1bii\_hispanic,model1bii\_asian,model1bii\_na,model1bii\_pi,model1bii\_2plus,model1bii\_other, nrow=4, top="Most minority groups have lower median housing prices compared to white populations")

# 1c. Is there a relationship between education and access?

model1c <- lm(ldist ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)

summary(model1c)

print("1c. The model is significant and has an adjusted R-squared of 4.446%. All of the coefficients are positive (decreased access compared to less than High School) except Bachelors and Doctorate, and most are significant except Masters and Doctorate. Notably, professional degree holders have significantly decreased access compared to other education groups.")

model1c\_HSdiploma <- ggplot(model1c$model, aes\_string(x = names(model1c$model)[2], y = names(model1c$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model1c)[1], slope = coef(model1c)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1c\_someCollege <- ggplot(model1c$model, aes\_string(x = names(model1c$model)[3], y = names(model1c$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1c)[1], slope = coef(model1c)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1c\_Associates <- ggplot(model1c$model, aes\_string(x = names(model1c$model)[4], y = names(model1c$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1c)[1], slope = coef(model1c)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1c\_Bachelors <- ggplot(model1c$model, aes\_string(x = names(model1c$model)[5], y = names(model1c$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1c)[1], slope = coef(model1c)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1c\_Masters <- ggplot(model1c$model, aes\_string(x = names(model1c$model)[6], y = names(model1c$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1c)[1], slope = coef(model1c)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1c\_Professional <- ggplot(model1c$model, aes\_string(x = names(model1c$model)[7], y = names(model1c$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model1c)[1], slope = coef(model1c)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1c\_Doctorate <- ggplot(model1c$model, aes\_string(x = names(model1c$model)[8], y = names(model1c$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model1c)[1], slope = coef(model1c)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1c\_HSdiploma,model1c\_someCollege,model1c\_Associates,model1c\_Bachelors,model1c\_Masters,model1c\_Professional,model1c\_Doctorate, nrow=4, top="Most education levels have decreased access to protected land")

# 1ci. Is there a relationship between education and income?

model1ci <- lm(average\_income ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)

summary(model1ci)

print("1c. The model is significant and has an adjusted R-squared of 78.06%. Most of the coefficients are significant except Associates, with all showing an increase in income (compared to less than High School) except Doctorate.")

model1ci\_HSdiploma <- ggplot(model1ci$model, aes\_string(x = names(model1ci$model)[2], y = names(model1ci$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model1ci)[1], slope = coef(model1ci)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1ci\_someCollege <- ggplot(model1ci$model, aes\_string(x = names(model1ci$model)[3], y = names(model1ci$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1ci)[1], slope = coef(model1ci)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1ci\_Associates <- ggplot(model1ci$model, aes\_string(x = names(model1ci$model)[4], y = names(model1ci$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1ci)[1], slope = coef(model1ci)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1ci\_Bachelors <- ggplot(model1ci$model, aes\_string(x = names(model1ci$model)[5], y = names(model1ci$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1ci)[1], slope = coef(model1ci)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1ci\_Masters <- ggplot(model1ci$model, aes\_string(x = names(model1ci$model)[6], y = names(model1ci$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1ci)[1], slope = coef(model1ci)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1ci\_Professional <- ggplot(model1ci$model, aes\_string(x = names(model1ci$model)[7], y = names(model1ci$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model1ci)[1], slope = coef(model1ci)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

model1ci\_Doctorate <- ggplot(model1ci$model, aes\_string(x = names(model1ci$model)[8], y = names(model1ci$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model1ci)[1], slope = coef(model1ci)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("avg\_income")

grid.arrange(model1ci\_HSdiploma,model1ci\_someCollege,model1ci\_Associates,model1ci\_Bachelors,model1ci\_Masters,model1ci\_Professional,model1ci\_Doctorate, nrow=4, top="Most education levels show increased average income compared to less than HS education")

# 1cii. Is there a relationship between education and housing price?

model1cii <- lm(median\_housing\_price ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)

summary(model1cii)

print("1c. The model is significant and has an adjusted R-squared of 65.71%. All coefficients are significant except Doctorate, which has a p-value of 0.2902. Compared to less the High School, Bachelors, Masters and Professional dregrees see increases in median housing price, while high school diploma, some college, Associates, and Doctorate show decreases in median housing price.")

model1cii\_HSdiploma <- ggplot(model1cii$model, aes\_string(x = names(model1cii$model)[2], y = names(model1cii$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model1cii)[1], slope = coef(model1cii)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1cii\_someCollege <- ggplot(model1cii$model, aes\_string(x = names(model1cii$model)[3], y = names(model1cii$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1cii)[1], slope = coef(model1cii)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1cii\_Associates <- ggplot(model1cii$model, aes\_string(x = names(model1cii$model)[4], y = names(model1cii$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1cii)[1], slope = coef(model1cii)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1cii\_Bachelors <- ggplot(model1cii$model, aes\_string(x = names(model1cii$model)[5], y = names(model1cii$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1cii)[1], slope = coef(model1cii)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1cii\_Masters <- ggplot(model1cii$model, aes\_string(x = names(model1cii$model)[6], y = names(model1cii$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1cii)[1], slope = coef(model1cii)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1cii\_Professional <- ggplot(model1cii$model, aes\_string(x = names(model1cii$model)[7], y = names(model1cii$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model1cii)[1], slope = coef(model1cii)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

model1cii\_Doctorate <- ggplot(model1cii$model, aes\_string(x = names(model1cii$model)[8], y = names(model1cii$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model1cii)[1], slope = coef(model1cii)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1)) + ylab("med\_hse\_pr")

grid.arrange(model1cii\_HSdiploma,model1cii\_someCollege,model1cii\_Associates,model1cii\_Bachelors,model1cii\_Masters,model1cii\_Professional,model1cii\_Doctorate, nrow=4, top="Bachelors, Masters and Professional levels of education show increased median housing prices")

# 1d. What are primary indicators of access, all else constant?

model1d <- lm(ldist ~ average\_income + median\_housing\_price + urban + suburb + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)

summary(model1d)

print("1d. The model is significant and has an adjusted R-squared of 15.88%. The significant variables are average income, median housing price, urban, suburban, all race/ethnicity variables except other, high school diploma, Bachelors, and Doctorate degrees.")

model1d\_average\_income <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[2], y = names(model1d$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[2], colour="black", size=1)

model1d\_median\_housing\_price <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[3], y = names(model1d$model)[1])) + geom\_point(colour="coral", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[3], colour="black", size=1)

model1d\_urban <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[4], y = names(model1d$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_suburb <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[5], y = names(model1d$model)[1])) + geom\_point(colour="chartreuse4", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_black <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[6], y = names(model1d$model)[1])) + geom\_point(colour="cadetblue", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_hispanic <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[7], y = names(model1d$model)[1])) + geom\_point(colour="burlywood", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_asian <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[8], y = names(model1d$model)[1])) + geom\_point(colour="brown2", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_na <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[9], y = names(model1d$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[9], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_pi <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[10], y = names(model1d$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[10], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_2plus <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[11], y = names(model1d$model)[1])) + geom\_point(colour="bisque3", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[11], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_other <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[12], y = names(model1d$model)[1])) + geom\_point(colour="aquamarine3", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[12], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_HSdiploma <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[13], y = names(model1d$model)[1])) + geom\_point(colour="dodgerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[13], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_someCollege <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[14], y = names(model1d$model)[1])) + geom\_point(colour="goldenrod1", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[14], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_Associates <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[15], y = names(model1d$model)[1])) + geom\_point(colour="forestgreen", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[15], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_Bachelors <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[16], y = names(model1d$model)[1])) + geom\_point(colour="indianred1", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[16], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_Masters <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[17], y = names(model1d$model)[1])) + geom\_point(colour="khaki", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[17], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_Professional <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[18], y = names(model1d$model)[1])) + geom\_point(colour="tomato", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[18], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1d\_Doctorate <- ggplot(model1d$model, aes\_string(x = names(model1d$model)[19], y = names(model1d$model)[1])) + geom\_point(colour="thistle", alpha = 0.1) + geom\_abline(intercept = coef(model1d)[1], slope = coef(model1d)[19], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1d\_average\_income, model1d\_median\_housing\_price, model1d\_urban, model1d\_suburb, model1d\_black, model1d\_hispanic, model1d\_asian, model1d\_na, model1d\_pi, model1d\_2plus, model1d\_other, model1d\_HSdiploma, model1d\_someCollege, model1d\_Associates, model1d\_Bachelors, model1d\_Masters, model1d\_Professional, model1d\_Doctorate, nrow=4, top="Access to protected land with all variables accounted for")

model1dRace <- lm(ldist ~ average\_income + median\_housing\_price + urban + suburb + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLand)

summary(model1dRace)

model1dRace\_average\_income <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[2], y = names(model1dRace$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[2], colour="black", size=1)

model1dRace\_median\_housing\_price <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[3], y = names(model1dRace$model)[1])) + geom\_point(colour="coral", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[3], colour="black", size=1)

model1dRace\_urban <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[4], y = names(model1dRace$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRace\_suburb <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[5], y = names(model1dRace$model)[1])) + geom\_point(colour="chartreuse4", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRace\_black <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[6], y = names(model1dRace$model)[1])) + geom\_point(colour="cadetblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRace\_hispanic <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[7], y = names(model1dRace$model)[1])) + geom\_point(colour="burlywood", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRace\_asian <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[8], y = names(model1dRace$model)[1])) + geom\_point(colour="brown2", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRace\_na <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[9], y = names(model1dRace$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[9], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRace\_pi <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[10], y = names(model1dRace$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[10], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRace\_2plus <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[11], y = names(model1dRace$model)[1])) + geom\_point(colour="bisque3", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[11], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRace\_other <- ggplot(model1dRace$model, aes\_string(x = names(model1dRace$model)[12], y = names(model1dRace$model)[1])) + geom\_point(colour="aquamarine3", alpha = 0.1) + geom\_abline(intercept = coef(model1dRace)[1], slope = coef(model1dRace)[12], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1dRace\_average\_income, model1dRace\_median\_housing\_price, model1dRace\_urban, model1dRace\_suburb, model1dRace\_black, model1dRace\_hispanic, model1dRace\_asian, model1dRace\_na, model1dRace\_pi, model1dRace\_2plus, model1dRace\_other, nrow=4, top="Access to protected land with all variables accounted for except education level")

model1dEdu <- lm(ldist ~ average\_income + median\_housing\_price + urban + suburb + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLand)

summary(model1dEdu)

model1dEdu\_average\_income <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[2], y = names(model1dEdu$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[2], colour="black", size=1)

model1dEdu\_median\_housing\_price <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[3], y = names(model1dEdu$model)[1])) + geom\_point(colour="coral", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[3], colour="black", size=1)

model1dEdu\_urban <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[4], y = names(model1dEdu$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dEdu\_suburb <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[5], y = names(model1dEdu$model)[1])) + geom\_point(colour="chartreuse4", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dEdu\_HSdiploma <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[6], y = names(model1dEdu$model)[1])) + geom\_point(colour="dodgerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dEdu\_someCollege <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[7], y = names(model1dEdu$model)[1])) + geom\_point(colour="goldenrod1", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dEdu\_Associates <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[8], y = names(model1dEdu$model)[1])) + geom\_point(colour="forestgreen", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dEdu\_Bachelors <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[9], y = names(model1dEdu$model)[1])) + geom\_point(colour="indianred1", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[9], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dEdu\_Masters <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[10], y = names(model1dEdu$model)[1])) + geom\_point(colour="khaki", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[10], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dEdu\_Professional <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[11], y = names(model1dEdu$model)[1])) + geom\_point(colour="tomato", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[11], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dEdu\_Doctorate <- ggplot(model1dEdu$model, aes\_string(x = names(model1dEdu$model)[12], y = names(model1dEdu$model)[1])) + geom\_point(colour="thistle", alpha = 0.1) + geom\_abline(intercept = coef(model1dEdu)[1], slope = coef(model1dEdu)[12], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1dEdu\_average\_income, model1dEdu\_median\_housing\_price, model1dEdu\_urban, model1dEdu\_suburb, model1dEdu\_HSdiploma, model1dEdu\_someCollege, model1dEdu\_Associates, model1dEdu\_Bachelors, model1dEdu\_Masters, model1dEdu\_Professional, model1dEdu\_Doctorate, nrow=4, top="Access to protected land with all variables accounted for except race")

print("What if we only look at urban/suburban areas, where unused green land areas--whether protected or not--are more scarce?")

ProtectedLandNonRural <- ProtectedLand[ProtectedLand$rural != 1,]

ProtectedLandUrban <- ProtectedLand[ProtectedLand$urban == 1,]

ProtectedLandSuburban <- ProtectedLand[ProtectedLand$suburb == 1,]

ProtectedLandRural <- ProtectedLand[ProtectedLand$rural == 1,]

model1dNonRural <- lm(ldist ~ average\_income + median\_housing\_price + urban + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandNonRural)

summary(model1dNonRural)

model1dNonRural\_average\_income <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[2], y = names(model1dNonRural$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[2], colour="black", size=1)

model1dNonRural\_median\_housing\_price <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[3], y = names(model1dNonRural$model)[1])) + geom\_point(colour="coral", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[3], colour="black", size=1)

model1dNonRural\_urban <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[4], y = names(model1dNonRural$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_black <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[5], y = names(model1dNonRural$model)[1])) + geom\_point(colour="cadetblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_hispanic <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[6], y = names(model1dNonRural$model)[1])) + geom\_point(colour="burlywood", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_asian <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[7], y = names(model1dNonRural$model)[1])) + geom\_point(colour="brown2", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_na <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[8], y = names(model1dNonRural$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_pi <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[9], y = names(model1dNonRural$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[9], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_2plus <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[10], y = names(model1dNonRural$model)[1])) + geom\_point(colour="bisque3", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[10], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_other <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[11], y = names(model1dNonRural$model)[1])) + geom\_point(colour="aquamarine3", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[11], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_HSdiploma <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[12], y = names(model1dNonRural$model)[1])) + geom\_point(colour="dodgerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[12], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_someCollege <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[13], y = names(model1dNonRural$model)[1])) + geom\_point(colour="goldenrod1", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[13], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_Associates <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[14], y = names(model1dNonRural$model)[1])) + geom\_point(colour="forestgreen", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[14], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_Bachelors <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[15], y = names(model1dNonRural$model)[1])) + geom\_point(colour="indianred1", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[15], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_Masters <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[16], y = names(model1dNonRural$model)[1])) + geom\_point(colour="khaki", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[16], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_Professional <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[17], y = names(model1dNonRural$model)[1])) + geom\_point(colour="tomato", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[17], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dNonRural\_Doctorate <- ggplot(model1dNonRural$model, aes\_string(x = names(model1dNonRural$model)[18], y = names(model1dNonRural$model)[1])) + geom\_point(colour="thistle", alpha = 0.1) + geom\_abline(intercept = coef(model1dNonRural)[1], slope = coef(model1dNonRural)[18], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1dNonRural\_average\_income, model1dNonRural\_median\_housing\_price, model1dNonRural\_urban, model1dNonRural\_black, model1dNonRural\_hispanic, model1dNonRural\_asian, model1dNonRural\_na, model1dNonRural\_pi, model1dNonRural\_2plus, model1dNonRural\_other, model1dNonRural\_HSdiploma, model1dNonRural\_someCollege, model1dNonRural\_Associates, model1dNonRural\_Bachelors, model1dNonRural\_Masters, model1dNonRural\_Professional, model1dNonRural\_Doctorate, nrow=4, top="Access to protected land with all variables accounted for except rural areas")

model1dUrban <- lm(ldist ~ average\_income + median\_housing\_price + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandUrban)

summary(model1dUrban)

model1dUrban\_average\_income <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[2], y = names(model1dUrban$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[2], colour="black", size=1)

model1dUrban\_median\_housing\_price <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[3], y = names(model1dUrban$model)[1])) + geom\_point(colour="coral", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[3], colour="black", size=1)

model1dUrban\_black <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[4], y = names(model1dUrban$model)[1])) + geom\_point(colour="cadetblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_hispanic <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[5], y = names(model1dUrban$model)[1])) + geom\_point(colour="burlywood", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_asian <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[6], y = names(model1dUrban$model)[1])) + geom\_point(colour="brown2", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_na <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[7], y = names(model1dUrban$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_pi <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[8], y = names(model1dUrban$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_2plus <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[9], y = names(model1dUrban$model)[1])) + geom\_point(colour="bisque3", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[9], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_other <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[10], y = names(model1dUrban$model)[1])) + geom\_point(colour="aquamarine3", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[10], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_HSdiploma <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[11], y = names(model1dUrban$model)[1])) + geom\_point(colour="dodgerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[11], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_someCollege <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[12], y = names(model1dUrban$model)[1])) + geom\_point(colour="goldenrod1", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[12], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_Associates <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[13], y = names(model1dUrban$model)[1])) + geom\_point(colour="forestgreen", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[13], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_Bachelors <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[14], y = names(model1dUrban$model)[1])) + geom\_point(colour="indianred1", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[14], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_Masters <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[15], y = names(model1dUrban$model)[1])) + geom\_point(colour="khaki", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[15], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_Professional <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[16], y = names(model1dUrban$model)[1])) + geom\_point(colour="tomato", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[16], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dUrban\_Doctorate <- ggplot(model1dUrban$model, aes\_string(x = names(model1dUrban$model)[17], y = names(model1dUrban$model)[1])) + geom\_point(colour="thistle", alpha = 0.1) + geom\_abline(intercept = coef(model1dUrban)[1], slope = coef(model1dUrban)[17], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1dUrban\_average\_income, model1dUrban\_median\_housing\_price, model1dUrban\_black, model1dUrban\_hispanic, model1dUrban\_asian, model1dUrban\_na, model1dUrban\_pi, model1dUrban\_2plus, model1dUrban\_other, model1dUrban\_HSdiploma, model1dUrban\_someCollege, model1dUrban\_Associates, model1dUrban\_Bachelors, model1dUrban\_Masters, model1dUrban\_Professional, model1dUrban\_Doctorate, nrow=4, top="Access to protected land with all variables accounted for in urban areas")

model1dSuburban <- lm(ldist ~ average\_income + median\_housing\_price + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandSuburban)

summary(model1dSuburban)

model1dSuburban\_average\_income <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[2], y = names(model1dSuburban$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[2], colour="black", size=1)

model1dSuburban\_median\_housing\_price <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[3], y = names(model1dSuburban$model)[1])) + geom\_point(colour="coral", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[3], colour="black", size=1)

model1dSuburban\_black <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[4], y = names(model1dSuburban$model)[1])) + geom\_point(colour="cadetblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_hispanic <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[5], y = names(model1dSuburban$model)[1])) + geom\_point(colour="burlywood", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_asian <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[6], y = names(model1dSuburban$model)[1])) + geom\_point(colour="brown2", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_na <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[7], y = names(model1dSuburban$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_pi <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[8], y = names(model1dSuburban$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_2plus <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[9], y = names(model1dSuburban$model)[1])) + geom\_point(colour="bisque3", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[9], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_other <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[10], y = names(model1dSuburban$model)[1])) + geom\_point(colour="aquamarine3", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[10], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_HSdiploma <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[11], y = names(model1dSuburban$model)[1])) + geom\_point(colour="dodgerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[11], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_someCollege <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[12], y = names(model1dSuburban$model)[1])) + geom\_point(colour="goldenrod1", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[12], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_Associates <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[13], y = names(model1dSuburban$model)[1])) + geom\_point(colour="forestgreen", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[13], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_Bachelors <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[14], y = names(model1dSuburban$model)[1])) + geom\_point(colour="indianred1", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[14], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_Masters <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[15], y = names(model1dSuburban$model)[1])) + geom\_point(colour="khaki", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[15], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_Professional <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[16], y = names(model1dSuburban$model)[1])) + geom\_point(colour="tomato", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[16], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dSuburban\_Doctorate <- ggplot(model1dSuburban$model, aes\_string(x = names(model1dSuburban$model)[17], y = names(model1dSuburban$model)[1])) + geom\_point(colour="thistle", alpha = 0.1) + geom\_abline(intercept = coef(model1dSuburban)[1], slope = coef(model1dSuburban)[17], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1dSuburban\_average\_income, model1dSuburban\_median\_housing\_price, model1dSuburban\_black, model1dSuburban\_hispanic, model1dSuburban\_asian, model1dSuburban\_na, model1dSuburban\_pi, model1dSuburban\_2plus, model1dSuburban\_other, model1dSuburban\_HSdiploma, model1dSuburban\_someCollege, model1dSuburban\_Associates, model1dSuburban\_Bachelors, model1dSuburban\_Masters, model1dSuburban\_Professional, model1dSuburban\_Doctorate, nrow=4, top="Access to protected land with all variables accounted for in suburban areas")

model1dRural <- lm(ldist ~ average\_income + median\_housing\_price + share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other + HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandRural)

summary(model1dRural)

model1dRural\_average\_income <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[2], y = names(model1dRural$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[2], colour="black", size=1)

model1dRural\_median\_housing\_price <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[3], y = names(model1dRural$model)[1])) + geom\_point(colour="coral", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[3], colour="black", size=1)

model1dRural\_black <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[4], y = names(model1dRural$model)[1])) + geom\_point(colour="cadetblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_hispanic <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[5], y = names(model1dRural$model)[1])) + geom\_point(colour="burlywood", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_asian <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[6], y = names(model1dRural$model)[1])) + geom\_point(colour="brown2", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_na <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[7], y = names(model1dRural$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_pi <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[8], y = names(model1dRural$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_2plus <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[9], y = names(model1dRural$model)[1])) + geom\_point(colour="bisque3", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[9], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_other <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[10], y = names(model1dRural$model)[1])) + geom\_point(colour="aquamarine3", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[10], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_HSdiploma <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[11], y = names(model1dRural$model)[1])) + geom\_point(colour="dodgerblue", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[11], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_someCollege <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[12], y = names(model1dRural$model)[1])) + geom\_point(colour="goldenrod1", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[12], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_Associates <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[13], y = names(model1dRural$model)[1])) + geom\_point(colour="forestgreen", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[13], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_Bachelors <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[14], y = names(model1dRural$model)[1])) + geom\_point(colour="indianred1", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[14], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_Masters <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[15], y = names(model1dRural$model)[1])) + geom\_point(colour="khaki", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[15], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_Professional <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[16], y = names(model1dRural$model)[1])) + geom\_point(colour="tomato", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[16], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model1dRural\_Doctorate <- ggplot(model1dRural$model, aes\_string(x = names(model1dRural$model)[17], y = names(model1dRural$model)[1])) + geom\_point(colour="thistle", alpha = 0.1) + geom\_abline(intercept = coef(model1dRural)[1], slope = coef(model1dRural)[17], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model1dRural\_average\_income, model1dRural\_median\_housing\_price, model1dRural\_black, model1dRural\_hispanic, model1dRural\_asian, model1dRural\_na, model1dRural\_pi, model1dRural\_2plus, model1dRural\_other, model1dRural\_HSdiploma, model1dRural\_someCollege, model1dRural\_Associates, model1dRural\_Bachelors, model1dRural\_Masters, model1dRural\_Professional, model1dRural\_Doctorate, nrow=4, top="Access to protected land with all variables accounted for in rural areas")

print("Looking at non-rural areas (urban and suburban combined), we see that a combination of median housing price and average income, when taken with appropriate magnitudes, would oftentimes indicate increased access for higher-income communities (overall negative effect on distance) over lower-income communities. We also see mostly negative coefficients (decreased distance and thus increased access) for more educated communities, compared to mostly positive coefficients (increased distance and thus decreased access) for less educated communities. Therefore, while we cannot establish racial inequality in access to protected lands in urban and suburban areas, we can reasonably state that there is income and education-related inequality.")

# 2. If there is inequality, what might reduce it?

print("2. Since protected land areas are commonly established through local and state government groups, departments, or officials, we will investigate the relationship between voter participation, party affiliation, and protected land areas.")

ProtectedLand$registration <- ProtectedLand$Registered/ProtectedLand$voting\_age\_pop

ProtectedLand$participation <- ProtectedLand$Voted/ProtectedLand$voting\_age\_pop

ProtectedLandNonRural$registration <- ProtectedLandNonRural$Registered/ProtectedLandNonRural$voting\_age\_pop

ProtectedLandNonRural$participation <- ProtectedLandNonRural$Voted/ProtectedLandNonRural$voting\_age\_pop

# 2a. Does access increase with voter registration?

model2a <- lm(ldist ~ registration, ProtectedLandNonRural)

summary(model2a)

print("2a. The coefficient indicates that in non-rural California, as voter registration (proportion of voting age population that is registered to vote) increases, distance to protected land decreases (increased access).")

ggplot(ProtectedLandNonRural, aes(x=registration,y=ldist)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + labs(title="Distance to protected land decreases as voter registration increases") + theme(plot.title = element\_text(hjust = 0.5), panel.grid.major=element\_blank(), panel.grid.minor=element\_blank()) + annotate("text", x = 1.5, y = 2.5, label = paste("Slope =", round(coef(model2a)[2],5), "\nP =", round(summary(model2a)$coef[2,4],4), "\nAdj. R2 = ", round(summary(model2a)$adj.r.squared,7)), colour="red")

# 2b. Does access increase with voter participation?

model2b <- lm(ldist ~ participation, ProtectedLandNonRural)

summary(model2b)

print("2b. This coefficient indicates that in non-rural California, as voter participation (proportion of voting age population that voted in the last election) increases, distance to protected land also increases (decreased access).")

ggplot(ProtectedLandNonRural, aes(x=participation,y=ldist)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + labs(title="Distance to protected land increases as voter participation increases") + theme(plot.title = element\_text(hjust = 0.5), panel.grid.major=element\_blank(), panel.grid.minor=element\_blank()) + annotate("text", x = 1, y = 2.5, label = paste("Slope =", round(coef(model2b)[2],5), "\nP =", round(summary(model2b)$coef[2,4],4), "\nAdj. R2 = ", round(summary(model2b)$adj.r.squared,7)), colour="red")

model2bi <- lm(ldist ~ registration + participation, ProtectedLandNonRural)

summary(model2bi)

print("This model shows the same coefficient signs from previous models at high levels of significance, indicating that higher voter registration and lower voter participation would be associated with increased access to protected land areas. Speculatively, this may indicate that voter engagement (interacting with local officials and politics in ways other than voting) may impact local protected areas more so than voter just participation.")

model2bi\_registration <- ggplot(model2bi$model, aes\_string(x = names(model2bi$model)[2], y = names(model2bi$model)[1])) + geom\_point(colour="gold", alpha = 0.05) + geom\_abline(intercept = coef(model2bi)[1], slope = coef(model2bi)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,2))

model2bi\_participation <- ggplot(model2bi$model, aes\_string(x = names(model2bi$model)[3], y = names(model2bi$model)[1])) + geom\_point(colour="skyblue", alpha = 0.05) + geom\_abline(intercept = coef(model2bi)[1], slope = coef(model2bi)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,2))

grid.arrange(model2bi\_registration,model2bi\_participation,nrow=2, top="Increased registration and less participation show an decreased distance to protected land")

# 2c. Does access depend on county-wide party affiliation?

model2c <- lm(ldist ~ Ratio.Dem.Rep, ProtectedLand)

summary(model2c)

print("2c. The variable Ratio.Dem.Rep represents the ratio of the proportion of registered Democrats to the proportion of registered Republicans at the county level. This model indicates that as the ratio increases by county (more registered Democrats compared to registered Republicans), the distance to protected land decreases for census tracts in that county.")

ggplot(ProtectedLand, aes(x=Ratio.Dem.Rep,y=ldist)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + labs(title="Distance to protected land decreases with increased registered Democrats") + theme(plot.title = element\_text(hjust = 0.5), panel.grid.major=element\_blank(), panel.grid.minor=element\_blank()) + annotate("text", x = 4.5, y = 2.5, label = paste("Slope =", round(coef(model2c)[2],5), "\nP =", round(summary(model2c)$coef[2,4],4), "\nAdj. R2 = ", round(summary(model2c)$adj.r.squared,7)), colour="red")

# 2ci. If there is a relationship, does it still exist when considering population density classification (rural, suburban, urban)?

model2ciNonRural <- lm(ldist ~ Ratio.Dem.Rep, ProtectedLandNonRural)

summary(model2ciNonRural)

model2ciUrban <- lm(ldist ~ Ratio.Dem.Rep, ProtectedLandUrban)

summary(model2ciUrban)

model2ciSuburban <- lm(ldist ~ Ratio.Dem.Rep, ProtectedLandSuburban)

summary(model2ciSuburban)

model2ciRural <- lm(ldist ~ Ratio.Dem.Rep, ProtectedLandRural)

summary(model2ciRural)

print("2ci. This relationship persists through all population density segments we have examined in this data set.")

model2ciNonRuralPlot <- ggplot(ProtectedLandNonRural, aes(x=Ratio.Dem.Rep,y=ldist)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + xlab("Ratio.Dem.Rep (Nonrural)")

model2ciUrbanPlot <- ggplot(ProtectedLandUrban, aes(x=Ratio.Dem.Rep,y=ldist)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + xlab("Ratio.Dem.Rep (Urban)")

model2ciSuburbanPlot <- ggplot(ProtectedLandSuburban, aes(x=Ratio.Dem.Rep,y=ldist)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + xlab("Ratio.Dem.Rep (Suburban)")

model2ciRuralPlot <- ggplot(model2ciRural, aes(x=Ratio.Dem.Rep,y=ldist)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + xlab("Ratio.Dem.Rep (Rural)")

grid.arrange(model2ciNonRuralPlot,model2ciUrbanPlot,model2ciSuburbanPlot,model2ciRuralPlot,nrow=2, top="Irrespective of population density classification, distance to protected land decreases with increased registered Democrats")

# 2d. Are there relationships between voter participation/registration and other demographic characteristics?

# 2di. Race/ethnicity?

model2diReg <- lm(registration ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLandNonRural)

summary(model2diReg)

model2diPart <- lm(participation ~ share\_black + share\_hispanic + share\_asian + share\_native\_american + share\_pacific\_islander + share\_2plus + share\_other, ProtectedLandNonRural)

summary(model2diPart)

print("2di. All minority populations show decreased voter registration and decreased voter participation compared to white populations in non-rural counties of California.")

model2diReg\_black <- ggplot(model2diReg$model, aes\_string(x = names(model2diReg$model)[2], y = names(model2diReg$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model2diReg)[1], slope = coef(model2diReg)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diReg\_hispanic <- ggplot(model2diReg$model, aes\_string(x = names(model2diReg$model)[3], y = names(model2diReg$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diReg)[1], slope = coef(model2diReg)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diReg\_asian <- ggplot(model2diReg$model, aes\_string(x = names(model2diReg$model)[4], y = names(model2diReg$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model2diReg)[1], slope = coef(model2diReg)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diReg\_na <- ggplot(model2diReg$model, aes\_string(x = names(model2diReg$model)[5], y = names(model2diReg$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diReg)[1], slope = coef(model2diReg)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diReg\_pi <- ggplot(model2diReg$model, aes\_string(x = names(model2diReg$model)[6], y = names(model2diReg$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model2diReg)[1], slope = coef(model2diReg)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diReg\_2plus <- ggplot(model2diReg$model, aes\_string(x = names(model2diReg$model)[7], y = names(model2diReg$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diReg)[1], slope = coef(model2diReg)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diReg\_other <- ggplot(model2diReg$model, aes\_string(x = names(model2diReg$model)[8], y = names(model2diReg$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model2diReg)[1], slope = coef(model2diReg)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model2diReg\_black, model2diReg\_hispanic, model2diReg\_asian, model2diReg\_na, model2diReg\_pi, model2diReg\_2plus, model2diReg\_other, nrow=4, top="All minority populations show decreased voter registration compared to white populations")

model2diPart\_black <- ggplot(model2diPart$model, aes\_string(x = names(model2diPart$model)[2], y = names(model2diPart$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model2diPart)[1], slope = coef(model2diPart)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diPart\_hispanic <- ggplot(model2diPart$model, aes\_string(x = names(model2diPart$model)[3], y = names(model2diPart$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diPart)[1], slope = coef(model2diPart)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diPart\_asian <- ggplot(model2diPart$model, aes\_string(x = names(model2diPart$model)[4], y = names(model2diPart$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model2diPart)[1], slope = coef(model2diPart)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diPart\_na <- ggplot(model2diPart$model, aes\_string(x = names(model2diPart$model)[5], y = names(model2diPart$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diPart)[1], slope = coef(model2diPart)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diPart\_pi <- ggplot(model2diPart$model, aes\_string(x = names(model2diPart$model)[6], y = names(model2diPart$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model2diPart)[1], slope = coef(model2diPart)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diPart\_2plus <- ggplot(model2diPart$model, aes\_string(x = names(model2diPart$model)[7], y = names(model2diPart$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diPart)[1], slope = coef(model2diPart)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diPart\_other <- ggplot(model2diPart$model, aes\_string(x = names(model2diPart$model)[8], y = names(model2diPart$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model2diPart)[1], slope = coef(model2diPart)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model2diPart\_black, model2diPart\_hispanic, model2diPart\_asian, model2diPart\_na, model2diPart\_pi, model2diPart\_2plus, model2diPart\_other, nrow=4, top="All minority populations show decreased voter participation compared to white populations")

# 2dii. Education?

model2diiReg <- lm(registration ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandNonRural)

summary(model2diiReg)

model2diiPart <- lm(participation ~ HSdiploma + someCollege + Associates + Bachelors + Masters + Professional + Doctorate, ProtectedLandNonRural)

summary(model2diiPart)

print("2dii. As education level increases, so does voter registration and voter turnout, except for communities with higher proportions of Doctorates, where we see turnout and participation levels much lower than all other education levels.")

model2diiReg\_HSdiploma <- ggplot(model2diiReg$model, aes\_string(x = names(model2diiReg$model)[2], y = names(model2diiReg$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model2diiReg)[1], slope = coef(model2diiReg)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiReg\_someCollege <- ggplot(model2diiReg$model, aes\_string(x = names(model2diiReg$model)[3], y = names(model2diiReg$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diiReg)[1], slope = coef(model2diiReg)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiReg\_Associates <- ggplot(model2diiReg$model, aes\_string(x = names(model2diiReg$model)[4], y = names(model2diiReg$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model2diiReg)[1], slope = coef(model2diiReg)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiReg\_Bachelors <- ggplot(model2diiReg$model, aes\_string(x = names(model2diiReg$model)[5], y = names(model2diiReg$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diiReg)[1], slope = coef(model2diiReg)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiReg\_Masters <- ggplot(model2diiReg$model, aes\_string(x = names(model2diiReg$model)[6], y = names(model2diiReg$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model2diiReg)[1], slope = coef(model2diiReg)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiReg\_Professional <- ggplot(model2diiReg$model, aes\_string(x = names(model2diiReg$model)[7], y = names(model2diiReg$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diiReg)[1], slope = coef(model2diiReg)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiReg\_Doctorate <- ggplot(model2diiReg$model, aes\_string(x = names(model2diiReg$model)[8], y = names(model2diiReg$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model2diiReg)[1], slope = coef(model2diiReg)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model2diiReg\_HSdiploma,model2diiReg\_someCollege,model2diiReg\_Associates,model2diiReg\_Bachelors,model2diiReg\_Masters,model2diiReg\_Professional,model2diiReg\_Doctorate, nrow=4, top="Higher levels of education typically lead to higher levels of voter registration")

model2diiPart\_HSdiploma <- ggplot(model2diiPart$model, aes\_string(x = names(model2diiPart$model)[2], y = names(model2diiPart$model)[1])) + geom\_point(colour="pink", alpha = 0.1) + geom\_abline(intercept = coef(model2diiPart)[1], slope = coef(model2diiPart)[2], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiPart\_someCollege <- ggplot(model2diiPart$model, aes\_string(x = names(model2diiPart$model)[3], y = names(model2diiPart$model)[1])) + geom\_point(colour="cornflowerblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diiPart)[1], slope = coef(model2diiPart)[3], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiPart\_Associates <- ggplot(model2diiPart$model, aes\_string(x = names(model2diiPart$model)[4], y = names(model2diiPart$model)[1])) + geom\_point(colour="chocolate", alpha = 0.1) + geom\_abline(intercept = coef(model2diiPart)[1], slope = coef(model2diiPart)[4], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiPart\_Bachelors <- ggplot(model2diiPart$model, aes\_string(x = names(model2diiPart$model)[5], y = names(model2diiPart$model)[1])) + geom\_point(colour="slateblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diiPart)[1], slope = coef(model2diiPart)[5], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiPart\_Masters <- ggplot(model2diiPart$model, aes\_string(x = names(model2diiPart$model)[6], y = names(model2diiPart$model)[1])) + geom\_point(colour="firebrick", alpha = 0.1) + geom\_abline(intercept = coef(model2diiPart)[1], slope = coef(model2diiPart)[6], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiPart\_Professional <- ggplot(model2diiPart$model, aes\_string(x = names(model2diiPart$model)[7], y = names(model2diiPart$model)[1])) + geom\_point(colour="lightblue", alpha = 0.1) + geom\_abline(intercept = coef(model2diiPart)[1], slope = coef(model2diiPart)[7], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

model2diiPart\_Doctorate <- ggplot(model2diiPart$model, aes\_string(x = names(model2diiPart$model)[8], y = names(model2diiPart$model)[1])) + geom\_point(colour="springgreen4", alpha = 0.1) + geom\_abline(intercept = coef(model2diiPart)[1], slope = coef(model2diiPart)[8], colour="black", size=1) + coord\_cartesian(xlim=c(0,1))

grid.arrange(model2diiPart\_HSdiploma,model2diiPart\_someCollege,model2diiPart\_Associates,model2diiPart\_Bachelors,model2diiPart\_Masters,model2diiPart\_Professional,model2diiPart\_Doctorate, nrow=4, top="Higher levels of education typically lead to higher levels of voter participation")

# 2diii. Income?

model2diiiReg <- lm(registration ~ average\_income, ProtectedLandNonRural)

summary(model2diiiReg)

model2diiiPart <- lm(participation ~ average\_income, ProtectedLandNonRural)

summary(model2diiiPart)

print("2diii. As income increases, so does voter registration and voter participation.")

ggplot(ProtectedLandNonRural, aes(x=average\_income,y=registration)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + labs(title="Increased average income shows increased voter registration") + theme(plot.title = element\_text(hjust = 0.5), panel.grid.major=element\_blank(), panel.grid.minor=element\_blank()) + annotate("text", x = 100000, y = 1.5, label = paste("Slope =", coef(model2diiiReg)[2], "\nP =", round(summary(model2diiiReg)$coef[2,4],4), "\nAdj. R2 = ", round(summary(model2diiiReg)$adj.r.squared,7)), colour="red")

ggplot(ProtectedLandNonRural, aes(x=average\_income,y=participation)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + labs(title="Increased average income shows increased voter participation") + theme(plot.title = element\_text(hjust = 0.5), panel.grid.major=element\_blank(), panel.grid.minor=element\_blank()) + annotate("text", x = 100000, y = 1.1, label = paste("Slope =", coef(model2diiiPart)[2], "\nP =", round(summary(model2diiiPart)$coef[2,4],4), "\nAdj. R2 = ", round(summary(model2diiiPart)$adj.r.squared,7)), colour="red")

# 2div. Housing price?

model2divReg <- lm(registration ~ median\_housing\_price, ProtectedLandNonRural)

summary(model2divReg)

model2divPart <- lm(participation ~ median\_housing\_price, ProtectedLandNonRural)

summary(model2divPart)

print("2div. As median housing price increases, so does voter registration and voter participation.")

ggplot(ProtectedLandNonRural, aes(x=median\_housing\_price,y=registration)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + labs(title="Increased median housing price shows increased voter registration") + theme(plot.title = element\_text(hjust = 0.5), panel.grid.major=element\_blank(), panel.grid.minor=element\_blank()) + annotate("text", x = 750000, y = 1.5, label = paste("Slope =", coef(model2divReg)[2], "\nP =", round(summary(model2divReg)$coef[2,4],4), "\nAdj. R2 = ", round(summary(model2divReg)$adj.r.squared,7)), colour="red")

ggplot(ProtectedLandNonRural, aes(x=median\_housing\_price,y=participation)) + geom\_point(alpha = 0.05) + geom\_smooth(method="lm") + labs(title="Increased median housing price shows increased voter participation") + theme(plot.title = element\_text(hjust = 0.5), panel.grid.major=element\_blank(), panel.grid.minor=element\_blank()) + annotate("text", x = 750000, y = 1, label = paste("Slope =", coef(model2divPart)[2], "\nP =", round(summary(model2divPart)$coef[2,4],4), "\nAdj. R2 = ", round(summary(model2divPart)$adj.r.squared,7)), colour="red")

print("Voter registration, a significant indicator of access to protected land areas by census tract, is reduced in certain communities of color, less educated communities, and lower income communities. Voter participation is similarly impacted in disadvantaged communities, but we do not see increased voter participation have the same association with access. With this knowledge, and without direct intervention to designate protected land areas in disadvantaged communities, we would suggest focusing on increasing voter registration in those communities as a gateway for increased voter engagement. That being said, direct intervention may be necessary, as disadvantaged communities often have more difficulty engaging with politics (lack of flexible working hours, inaccessibility to engagement opportunities).")

# 3. Can we use a small set of variables to identify non-rural census tracts where voter engagement (registration) is low and efforts to improve it would be most beneficial to the community?

model3Data <- na.omit(ProtectedLandNonRural)

randIndex <- sample(1:dim(model3Data)[1])

cutpoint <- floor(2\*dim(model3Data)[1]/3)

trainData <- model3Data[randIndex[1:cutpoint],]

testData <- model3Data[randIndex[(cutpoint+1):dim(model3Data)[1]],]

trainData$lowReg[trainData$registration < mean(model3Data$registration)] <- 1

trainData$lowReg[trainData$registration >= mean(model3Data$registration)] <- 0

testData$lowReg[testData$registration < mean(model3Data$registration)] <- 1

testData$lowReg[testData$registration >= mean(model3Data$registration)] <- 0

trainData$lowReg <- as.factor(trainData$lowReg)

testData$lowReg <- as.factor(testData$lowReg)

model3 <- ksvm(lowReg ~ median\_housing\_price + average\_income + lessthanHS + Ratio.Dem.Rep, data = trainData, kernel = "rbfdot", kpar = "automatic", C = 5, cross = 3, prob.model = TRUE)

model3

model3Pred <- predict(model3, testData)

testComp <- data.frame(testData, model3Pred)

correctPercent <- length(which(testComp$lowReg == testComp$model3Pred))/length(testComp$lowReg)

correctPercent

testComp$PredWrong[testComp$lowReg == testComp$model3Pred] <- 0

testComp$PredWrong[testComp$lowReg != testComp$model3Pred] <- 1

testComp$PredWrong <- as.factor(testComp$PredWrong)

model3plot <- ggplot(testComp) + geom\_point(aes(y=average\_income, x=lessthanHS, size=PredWrong, color=lowReg, shape=model3Pred)) + labs(title="Identifying non-rural areas where voter registration is low") + theme(plot.title = element\_text(hjust = 0.5))

model3plot

print("Using ksvm, we can use the data we have available to identify non-rural census tracts where voter engagement (registration) is low. Efforts to improve it would be most beneficial to the community and improve access to protected lands.")