DASC32103Project1-WIlliamBuckey

2025-02-05

```r
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(dplyr)
library(ggplot2)
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3

## corrplot 0.95 loaded
```

```r
library(ggpubr)
library(cowplot)
```

```
## Warning: package 'cowplot' was built under R version 4.3.2

##
## Attaching package: 'cowplot'
##
## The following object is masked from 'package:ggpubr':
##
##     get_legend
##
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```r
mobility_data <- read.csv("mobility-all.csv")
```

```r
# View basic structure
str(mobility_data)
```

```
## 'data.frame':    741 obs. of  43 variables:
##  $ ID                        : int  100 200 301 302 401 402 500 601 602 700 ...
##  $ Name                      : chr  "Johnson City" "Morristown" "Middlesborough" "Knoxville" ...
##  $ Mobility                  : num  0.0622 0.0537 0.0726 0.0563 0.0448 ...
##  $ State                     : chr  "TN" "TN" "TN" "TN" ...
```

```
##  $ Population                : int  576081 227816 66708 727600 493180 92753 1055133 90016 64676 35453
##  $ Urban                     : int  1 1 0 1 1 0 1 0 0 1 ...
##  $ Black                     : num  0.021 0.02 0.015 0.056 0.174 0.224 0.218 0.032 0.029 0.207 ...
##  $ Seg_racial                : num  0.09 0.093 0.064 0.21 0.262 0.137 0.22 0.114 0.131 0.139 ...
##  $ Seg_income                : num  0.035 0.026 0.024 0.092 0.072 0.024 0.068 0.012 0.005 0.045 ...
##  $ Seg_poverty               : num  0.03 0.028 0.015 0.084 0.061 0.015 0.058 0.009 0.004 0.044 ...
##  $ Seg_affluence             : num  0.038 0.025 0.026 0.102 0.081 0.028 0.077 0.012 0.006 0.045 ...
##  $ Commute                   : num  0.325 0.276 0.359 0.269 0.292 0.313 0.305 0.289 0.325 0.299 ...
##  $ Income                    : int  31560 29959 22328 35884 38892 31265 36582 31544 30683 33417 ...
##  $ Gini                      : num  0.468 0.435 0.441 0.508 0.466 0.444 0.524 0.446 0.356 0.471 ...
##  $ Share01                   : num  13.5 10.6 10.7 15.1 11.9 ...
##  $ Gini_99                   : num  0.333 0.328 0.334 0.358 0.346 0.338 0.341 0.32 0.269 0.341 ...
##  $ Middle_class              : num  0.548 0.538 0.467 0.504 0.5 0.538 0.51 0.56 0.608 0.529 ...
##  $ Local_tax_rate            : num  0.02 0.023 0.015 0.019 0.018 0.015 0.017 0.014 0.014 0.018 ...
##  $ Local_gov_spending        : int  1886 2004 1190 2357 1891 1558 1932 1661 1208 2499 ...
##  $ Progressivity             : num  0 0 0 0 1 0 1 1 0 0 ...
##  $ EITC                      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ School_spending           : num  5.18 4.51 5.61 4.9 5.46 ...
##  $ Student_teacher_ratio     : num  NA NA 15.1 NA 15.4 NA 16.7 16.2 12.3 15.9 ...
##  $ Test_scores               : num  2.73 -3.4 -9.31 -6.03 -2.3 ...
##  $ HS_dropout                : num  -0.015 -0.024 -0.005 -0.011 0.023 NA 0.016 0.021 NA NA ...
##  $ Colleges                  : num  0.014 0.009 0.045 0.011 0.014 0.011 0.014 0.011 NA 0.02 ...
##  $ Tuition                   : int  4817 4762 11840 3480 9715 1113 4528 880 NA 7264 ...
##  $ Graduation                : num  -0.002 -0.101 0.111 -0.024 0.052 -0.116 -0.017 -0.123 NA 0.007 ..
##  $ Labor_force_participation: num  0.587 0.625 0.479 0.615 0.656 0.599 0.666 0.617 0.594 0.63 ...
##  $ Manufacturing             : num  0.237 0.238 0.234 0.146 0.215 0.395 0.261 0.275 0.321 0.295 ...
##  $ Chinese_imports           : num  5.29 3.03 2.06 1.08 1.02 ...
##  $ Teenage_labor             : num  0.004 0.005 0.003 0.004 0.004 0.003 0.004 0.003 0.004 0.004 ...
##  $ Migration_in              : num  0.006 0.016 0.008 0.016 0.022 0.007 0.017 0.012 0.006 0.017 ...
##  $ Migration_out             : num  0.005 0.014 0.012 0.014 0.019 0.01 0.015 0.012 0.006 0.016 ...
##  $ Foreign_born              : num  0.012 0.023 0.007 0.02 0.053 0.025 0.05 0.027 0.023 0.029 ...
##  $ Social_capital            : num  -0.298 -0.767 -1.27 -0.222 -0.018 ...
##  $ Religious                 : num  0.514 0.544 0.668 0.602 0.488 0.454 0.434 0.561 0.43 0.596 ...
##  $ Violent_crime             : num  0.001 0.002 0.001 0.001 0.003 0.002 0.003 0.003 0.001 0.003 ...
##  $ Single_mothers            : num  0.19 0.185 0.211 0.206 0.22 0.241 0.237 0.165 0.167 0.246 ...
##  $ Divorced                  : num  0.11 0.116 0.113 0.114 0.092 0.096 0.096 0.087 0.089 0.099 ...
##  $ Married                   : num  0.601 0.613 0.59 0.575 0.586 0.58 0.56 0.632 0.622 0.561 ...
##  $ Longitude                 : num  -82.4 -83.4 -83.5 -84.2 -80.5 ...
##  $ Latitude                  : num  36.5 36.1 36.6 36 36.1 ...
```

```r
# Check for missing values
colSums(is.na(mobility_data))
```

```
##                        ID                      Name                  Mobility
##                         0                         0                        12
##                     State                Population                     Urban
##                         0                         0                         0
##                     Black                Seg_racial                Seg_income
##                         0                         0                         0
##               Seg_poverty             Seg_affluence                   Commute
##                         0                         0                         0
##                    Income                      Gini                   Share01
##                         0                         0                        32
##                   Gini_99              Middle_class            Local_tax_rate
##                        32                        32                         1
```

```
##        Local_gov_spending              Progressivity                       EITC
##                         2                          0                          0
##           School_spending      Student_teacher_ratio                Test_scores
##                        10                         30                         36
##                HS_dropout                   Colleges                    Tuition
##                       148                        157                        161
##                Graduation Labor_force_participation              Manufacturing
##                       160                          0                          0
##           Chinese_imports              Teenage_labor               Migration_in
##                        19                         32                         17
##             Migration_out               Foreign_born             Social_capital
##                        17                          0                         19
##                 Religious              Violent_crime             Single_mothers
##                         0                         27                          0
##                  Divorced                    Married                  Longitude
##                         0                          0                          0
##                  Latitude
##                         0
```

```r
# Convert categorical variables to factors
mobility_data$State <- as.factor(mobility_data$State)
mobility_data$Urban <- as.factor(mobility_data$Urban)  # If applicable

# Summary statistics for numerical variables
summary(mobility_data)
```

```
##        ID             Name              Mobility            State
##  Min.   :  100   Length:741         Min.   :0.02210   TX     : 64
##  1st Qu.:12701   Class :character   1st Qu.:0.06599   KS     : 32
##  Median :26106   Mode  :character   Median :0.08951   GA     : 28
##  Mean   :22444                      Mean   :0.10042   MO     : 24
##  3rd Qu.:31301                      3rd Qu.:0.11940   SD     : 24
##  Max.   :39400                      Max.   :0.46970   MN     : 23
##                                     NA's   :12        (Other):546
##    Population        Urban       Black             Seg_racial
##  Min.   :     1193   0:416   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:    38384   1:325   1st Qu.:0.00400   1st Qu.:0.0560
##  Median :   103842           Median :0.02200   Median :0.1070
##  Mean   :   379787           Mean   :0.07781   Mean   :0.1298
##  3rd Qu.:   289849           3rd Qu.:0.08200   3rd Qu.:0.1810
##  Max.   :16393360            Max.   :0.65800   Max.   :0.5540
##
##    Seg_income        Seg_poverty       Seg_affluence        Commute
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.1560
##  1st Qu.:0.01400   1st Qu.:0.01300   1st Qu.:0.01300   1st Qu.:0.3450
##  Median :0.03100   Median :0.02800   Median :0.03200   Median :0.4360
##  Mean   :0.03952   Mean   :0.03626   Mean   :0.04162   Mean   :0.4572
##  3rd Qu.:0.05700   3rd Qu.:0.05400   3rd Qu.:0.06000   3rd Qu.:0.5630
##  Max.   :0.13800   Max.   :0.12900   Max.   :0.15400   Max.   :0.9450
##
##      Income           Gini           Share01          Gini_99
##  Min.   :16696   Min.   :0.2020   Min.   : 2.673   Min.   :0.0630
##  1st Qu.:29327   1st Qu.:0.3480   1st Qu.: 8.005   1st Qu.:0.2570
##  Median :32372   Median :0.3980   Median :10.119   Median :0.2990
##  Mean   :32870   Mean   :0.4055   Mean   :10.842   Mean   :0.3012
```

```
##  3rd Qu.:35816   3rd Qu.:0.4570   3rd Qu.:12.545   3rd Qu.:0.3410
##  Max.   :58628   Max.   :0.8470   Max.   :64.788   Max.   :0.4470
##                                   NA's   :32       NA's   :32
##   Middle_class    Local_tax_rate    Local_gov_spending Progressivity
##  Min.   :0.2850   Min.   :0.00800   Min.   :  952      Min.   :0.0000
##  1st Qu.:0.5000   1st Qu.:0.01700   1st Qu.: 1722      1st Qu.:0.0000
##  Median :0.5520   Median :0.02200   Median : 2112      Median :0.0000
##  Mean   :0.5499   Mean   :0.02359   Mean   : 2309      Mean   :0.7738
##  3rd Qu.:0.6080   3rd Qu.:0.02700   3rd Qu.: 2638      3rd Qu.:1.0000
##  Max.   :0.7340   Max.   :0.08200   Max.   :13621      Max.   :7.2200
##  NA's   :32       NA's   :1         NA's   :2
##       EITC        School_spending  Student_teacher_ratio  Test_scores
##  Min.   : 0.000   Min.   : 3.920   Min.   : 9.60         Min.   :-32.78500
##  1st Qu.: 0.000   1st Qu.: 5.168   1st Qu.:14.90         1st Qu.: -4.29300
##  Median : 0.000   Median : 5.897   Median :16.50         Median :  0.74100
##  Mean   : 1.334   Mean   : 6.037   Mean   :16.51         Mean   :  0.00001
##  3rd Qu.: 0.000   3rd Qu.: 6.627   3rd Qu.:18.00         3rd Qu.:  5.55400
##  Max.   :21.333   Max.   :11.906   Max.   :27.70         Max.   : 20.07100
##                   NA's   :10       NA's   :30            NA's   :36
##    HS_dropout          Colleges          Tuition        Graduation
##  Min.   :-0.04300   Min.   :0.00100   Min.   :    0   Min.   :-0.35000
##  1st Qu.:-0.01500   1st Qu.:0.01200   1st Qu.: 1631   1st Qu.:-0.09700
##  Median :-0.00400   Median :0.01700   Median : 2938   Median :-0.01600
##  Mean   :-0.00001   Mean   :0.02311   Mean   : 4355   Mean   :-0.00001
##  3rd Qu.: 0.01100   3rd Qu.:0.02600   3rd Qu.: 5866   3rd Qu.: 0.08300
##  Max.   : 0.10900   Max.   :0.24300   Max.   :24619   Max.   : 0.52800
##  NA's   :148        NA's   :157       NA's   :161     NA's   :160
##  Labor_force_participation Manufacturing    Chinese_imports   Teenage_labor
##  Min.   :0.364             Min.   :0.0020   Min.   :-0.0750   Min.   :0.00200
##  1st Qu.:0.581             1st Qu.:0.0760   1st Qu.: 0.2605   1st Qu.:0.00400
##  Median :0.619             Median :0.1330   Median : 0.7455   Median :0.00500
##  Mean   :0.616             Mean   :0.1404   Mean   : 1.1757   Mean   :0.00483
##  3rd Qu.:0.653             3rd Qu.:0.1990   3rd Qu.: 1.4145   3rd Qu.:0.00600
##  Max.   :0.816             Max.   :0.4490   Max.   :25.4050   Max.   :0.00800
##                                             NA's   :19        NA's   :32
##   Migration_in      Migration_out      Foreign_born      Social_capital
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :-3.1990
##  1st Qu.:0.01000   1st Qu.:0.01200   1st Qu.:0.01200   1st Qu.:-0.7655
##  Median :0.01400   Median :0.01600   Median :0.02400   Median : 0.0640
##  Mean   :0.01653   Mean   :0.01683   Mean   :0.04117   Mean   : 0.1717
##  3rd Qu.:0.02100   3rd Qu.:0.02100   3rd Qu.:0.04600   3rd Qu.: 0.9653
##  Max.   :0.07700   Max.   :0.05200   Max.   :0.39700   Max.   : 7.3050
##  NA's   :17        NA's   :17                          NA's   :19
##    Religious      Violent_crime      Single_mothers      Divorced
##  Min.   :0.1100   Min.   :0.000000   Min.   :0.0820   Min.   :0.04000
##  1st Qu.:0.4250   1st Qu.:0.001000   1st Qu.:0.1710   1st Qu.:0.08500
##  Median :0.5250   Median :0.001000   Median :0.1960   Median :0.09800
##  Mean   :0.5456   Mean   :0.001594   Mean   :0.2017   Mean   :0.09666
##  3rd Qu.:0.6430   3rd Qu.:0.002000   3rd Qu.:0.2260   3rd Qu.:0.10900
##  Max.   :1.3080   Max.   :0.028000   Max.   :0.4340   Max.   :0.19000
##                   NA's   :27
##     Married         Longitude          Latitude
##  Min.   :0.3730   Min.   :-170.72   Min.   :19.58
##  1st Qu.:0.5450   1st Qu.:-101.53   1st Qu.:34.80
```

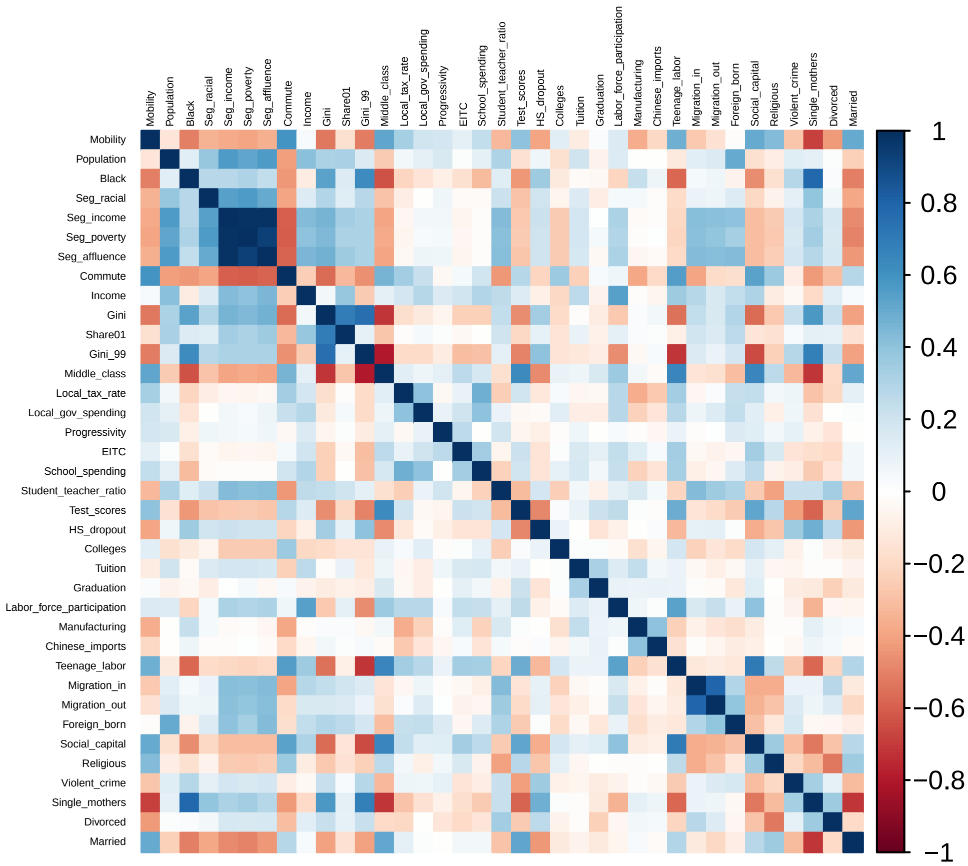
```
##  Median :0.5800   Median : -93.63   Median :38.92
##  Mean   :0.5745   Mean   : -95.55   Mean   :39.06
##  3rd Qu.:0.6070   3rd Qu.: -84.79   3rd Qu.:42.88
##  Max.   :0.6950   Max.   : -67.61   Max.   :68.37
##
```

```r
# Identify columns with missing values
missing_values <- colSums(is.na(mobility_data))
missing_values[missing_values > 0]  # Show only columns with missing values
```

```
##              Mobility               Share01               Gini_99
##                    12                    32                    32
##          Middle_class        Local_tax_rate    Local_gov_spending
##                    32                     1                     2
##       School_spending Student_teacher_ratio           Test_scores
##                    10                    30                    36
##            HS_dropout              Colleges               Tuition
##                   148                   157                   161
##            Graduation       Chinese_imports         Teenage_labor
##                   160                    19                    32
##          Migration_in         Migration_out        Social_capital
##                    17                    17                    19
##         Violent_crime
##                    27
```

```r
# Handle missing values (Options: Remove or Impute)
mobility_data <- mobility_data %>%
  drop_na(Mobility)  # Remove rows where Mobility is missing

# Alternatively, impute missing values using median for numerical variables
mobility_data <- mobility_data %>%
  mutate(across(where(is.numeric), ~ ifelse(is.na(.), median(., na.rm = TRUE), .)))

# Save cleaned dataset for future steps
write.csv(mobility_data, "mobility_cleaned.csv", row.names = FALSE)
```

```r
# Step 1: Remove non-informative numeric columns (ID, Longitude, Latitude)
cleaned_data <- mobility_data %>%
  select(-c(ID, Longitude, Latitude))

# Step 2: Remove columns with ANY missing values
cleaned_data <- mobility_data %>%
  select(where(is.numeric)) %>%  # Keeps only numeric variables
  select(-c(ID, Longitude, Latitude))  # Explicitly remove ID & coordinates

# Step 3: Compute correlation matrix
cor_matrix <- cor(cleaned_data, use = "pairwise.complete.obs")

# Step 4: Create the full heatmap (WITHOUT numbers)
corrplot(cor_matrix,
         method = "color",  # Color-coded correlation plot
         tl.col = "black",  # Black text labels
         tl.cex = 0.3)      # Adjust text size for readability
```




```r
cor_df <- as.data.frame(as.table(cor_matrix))

# Step 4: Remove self-correlations (diagonal)
cor_df <- cor_df %>%
  filter(Var1 != Var2)

# Step 5: Standardize Var1 & Var2 order to remove duplicates
cor_df <- cor_df %>%
  rowwise() %>%
  mutate(pair = paste(sort(c(Var1, Var2)), collapse = "_")) %>%  # Create a unique pair ID
  distinct(pair, .keep_all = TRUE) %>%  # Remove duplicate pairs
  select(-pair)

# Step 4: Sort by absolute correlation strength (highest to lowest)
top_corr <- cor_df %>%
  arrange(desc(abs(Freq))) %>%  # Sort by absolute correlation
  head(50)  # Select top 30

# Step 5: Print top 50 correlated variable pairs
print(top_corr)
```

```
## # A tibble: 50 x 3
## # Rowwise:
##    Var1          Var2            Freq
##    <fct>         <fct>          <dbl>
##  1 Seg_affluence Seg_income     0.986
##  2 Seg_poverty   Seg_income     0.981
```

```
##   3 Seg_affluence  Seg_poverty     0.939
##   4 Middle_class   Gini_99        -0.795
##   5 Migration_out  Migration_in    0.793
##   6 Single_mothers Black           0.781
##   7 Gini_99        Gini            0.753
##   8 Married        Single_mothers -0.716
##   9 Middle_class   Gini           -0.715
##  10 Teenage_labor  Gini_99        -0.715
## # i 40 more rows
```

```r
# Load required libraries
library(dplyr)

# Step 1: Define policy-driven variables
policy_vars <- c("Local_tax_rate", "Local_gov_spending", "Progressivity", "EITC",
                 "School_spending", "Student_teacher_ratio", "Test_scores",
                 "HS_dropout", "Labor_force_participation", "Social_capital",
                 "Colleges", "Tuition", "Single_mothers")

# Step 2: Compute correlation matrix
cor_matrix <- cor(cleaned_data, use = "pairwise.complete.obs")

# Step 3: Convert matrix into a dataframe
cor_df <- as.data.frame(as.table(cor_matrix))

# Step 4: Remove self-correlations (diagonal)
cor_df <- cor_df %>%
  filter(Var1 != Var2)

# Step 5: Standardize Var1 & Var2 order to remove duplicates
cor_df <- cor_df %>%
  rowwise() %>%
  mutate(pair = paste(sort(c(Var1, Var2)), collapse = "_")) %>%  # Create a unique pair ID
  distinct(pair, .keep_all = TRUE) %>%  # Remove duplicate pairs
  select(-pair)  # Drop helper column

# Step 6: Find top 5 correlated variables for each policy predictor
top_correlations <- list()

for (var in policy_vars) {
  top_5 <- cor_df %>%
    filter(Var1 == var | Var2 == var) %>%  # Select rows where var appears
    arrange(desc(abs(Freq))) %>%  # Sort by absolute correlation
    head(5)  # Select top 5
  top_correlations[[var]] <- top_5
}

# Step 7: Display results
print(top_correlations)
```

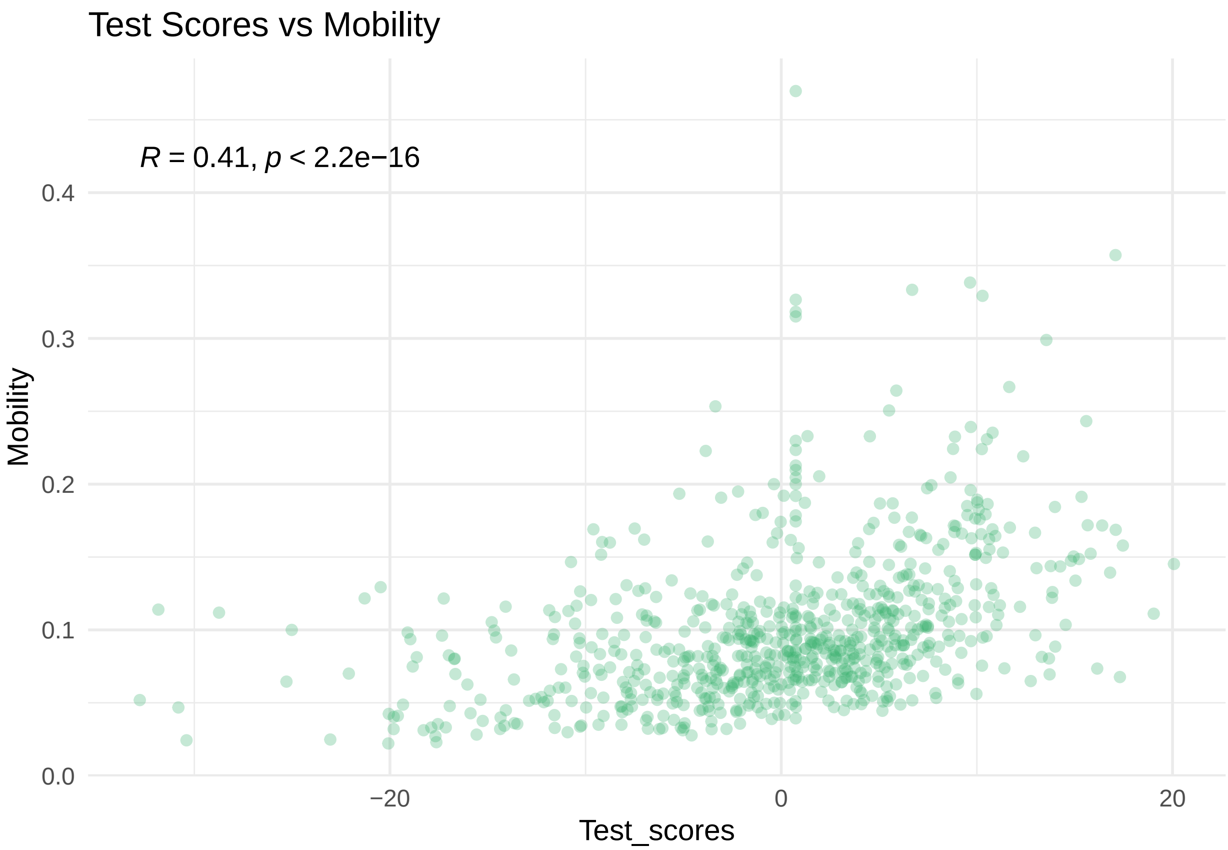
```
## $Local_tax_rate
## # A tibble: 5 x 3
## # Rowwise:
##    Var1                Var2             Freq
##    <fct>               <fct>           <dbl>
```

```
## 1 School_spending    Local_tax_rate  0.486
## 2 Local_gov_spending Local_tax_rate  0.406
## 3 Manufacturing      Local_tax_rate -0.362
## 4 Local_tax_rate     Commute         0.350
## 5 Teenage_labor      Local_tax_rate  0.349
##
## $Local_gov_spending
## # A tibble: 5 x 3
## # Rowwise:
##   Var1                     Var2                Freq
##   <fct>                    <fct>              <dbl>
## 1 Local_gov_spending       Local_tax_rate     0.406
## 2 School_spending          Local_gov_spending 0.403
## 3 Local_gov_spending       Income             0.285
## 4 Teenage_labor            Local_gov_spending 0.275
## 5 Labor_force_participation Local_gov_spending 0.271
##
## $Progressivity
## # A tibble: 5 x 3
## # Rowwise:
##   Var1                  Var2           Freq
##   <fct>                 <fct>         <dbl>
## 1 EITC                  Progressivity 0.262
## 2 Student_teacher_ratio Progressivity 0.197
## 3 Progressivity         Mobility      0.190
## 4 Progressivity         Population    0.160
## 5 Foreign_born          Progressivity 0.154
##
## $EITC
## # A tibble: 5 x 3
## # Rowwise:
##   Var1            Var2           Freq
##   <fct>           <fct>         <dbl>
## 1 Teenage_labor   EITC          0.350
## 2 School_spending EITC          0.349
## 3 Social_capital  EITC          0.345
## 4 EITC            Gini_99      -0.305
## 5 EITC            Middle_class  0.268
##
## $School_spending
## # A tibble: 5 x 3
## # Rowwise:
##   Var1            Var2                Freq
##   <fct>           <fct>              <dbl>
## 1 School_spending Local_tax_rate      0.486
## 2 School_spending Local_gov_spending  0.403
## 3 School_spending EITC                0.349
## 4 Teenage_labor   School_spending     0.335
## 5 School_spending Black              -0.311
##
## $Student_teacher_ratio
## # A tibble: 5 x 3
## # Rowwise:
##   Var1                  Var2                    Freq
```

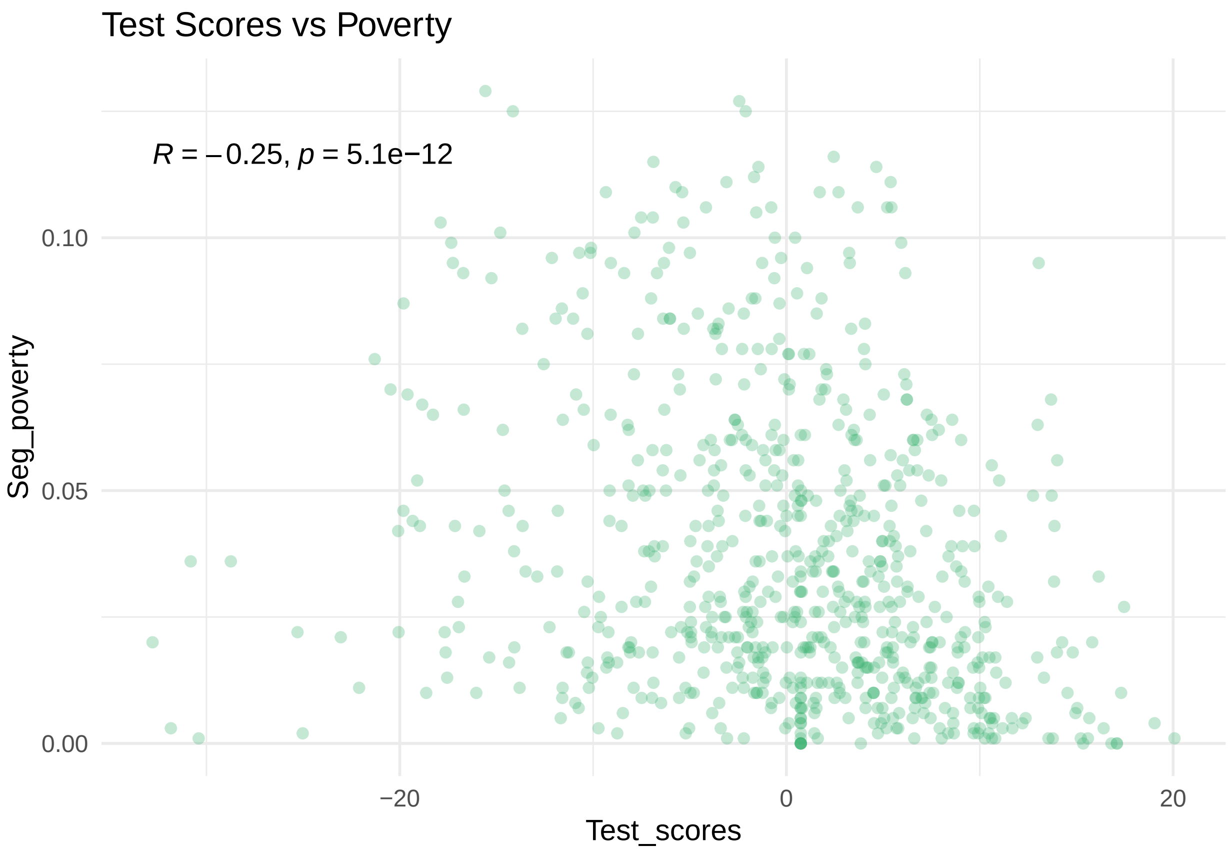
```
##   <fct>                 <fct>                  <dbl>
## 1 Migration_in          Student_teacher_ratio  0.435
## 2 Student_teacher_ratio Seg_income             0.432
## 3 Student_teacher_ratio Commute               -0.431
## 4 Student_teacher_ratio Seg_affluence          0.428
## 5 Student_teacher_ratio Seg_poverty            0.417
##
## $Test_scores
## # A tibble: 5 x 3
## # Rowwise:
##   Var1           Var2           Freq
##   <fct>          <fct>         <dbl>
## 1 Test_scores    Middle_class  0.638
## 2 Single_mothers Test_scores  -0.580
## 3 Social_capital Test_scores   0.523
## 4 Married        Test_scores   0.521
## 5 Test_scores    Gini_99      -0.496
##
## $HS_dropout
## # A tibble: 5 x 3
## # Rowwise:
##   Var1           Var2           Freq
##   <fct>          <fct>         <dbl>
## 1 HS_dropout     Test_scores  -0.487
## 2 Single_mothers HS_dropout    0.482
## 3 HS_dropout     Middle_class -0.474
## 4 Married        HS_dropout   -0.432
## 5 HS_dropout     Gini_99       0.402
##
## $Labor_force_participation
## # A tibble: 5 x 3
## # Rowwise:
##   Var1                      Var2                        Freq
##   <fct>                     <fct>                      <dbl>
## 1 Labor_force_participation Income                     0.544
## 2 Teenage_labor             Labor_force_participation  0.534
## 3 Labor_force_participation Gini_99                   -0.465
## 4 Social_capital            Labor_force_participation  0.403
## 5 Labor_force_participation Middle_class               0.361
##
## $Social_capital
## # A tibble: 5 x 3
## # Rowwise:
##   Var1           Var2            Freq
##   <fct>          <fct>          <dbl>
## 1 Social_capital Teenage_labor  0.708
## 2 Social_capital Gini_99       -0.656
## 3 Social_capital Middle_class   0.652
## 4 Social_capital Gini          -0.569
## 5 Social_capital Commute        0.531
##
## $Colleges
## # A tibble: 5 x 3
## # Rowwise:
```

```
##   Var1     Var2                    Freq
##   <fct>    <fct>                  <dbl>
## 1 Colleges Commute                 0.360
## 2 Colleges Seg_affluence          -0.260
## 3 Colleges Seg_income             -0.257
## 4 Colleges Seg_poverty            -0.251
## 5 Colleges Student_teacher_ratio -0.242
##
## $Tuition
## # A tibble: 5 x 3
## # Rowwise:
##   Var1          Var2         Freq
##   <fct>         <fct>       <dbl>
## 1 Graduation    Tuition     0.325
## 2 Tuition       Income      0.260
## 3 Manufacturing Tuition     0.244
## 4 Tuition       Commute    -0.231
## 5 Tuition       Population  0.203
##
## $Single_mothers
## # A tibble: 5 x 3
## # Rowwise:
##   Var1           Var2             Freq
##   <fct>          <fct>           <dbl>
## 1 Single_mothers Black           0.781
## 2 Married        Single_mothers -0.716
## 3 Single_mothers Middle_class   -0.711
## 4 Single_mothers Mobility       -0.686
## 5 Single_mothers Gini_99         0.683
```

```r
# Define base dataset
data <- cleaned_data  # Use cleaned dataset without missing values

# Function to create individual scatter plots (Fixed for ggplot2 3.0+)
plot_scatter <- function(x_var, y_var, color, title) {
  ggplot(data, aes(.data[[x_var]], .data[[y_var]])) +  # Updated for tidy evaluation
    geom_point(color = color, alpha = .3) +
    stat_cor(label.x = min(data[[x_var]], na.rm = TRUE),
             label.y = max(data[[y_var]], na.rm = TRUE) * 0.9) +
    ggtitle(title) +
    theme_minimal()
}

# Generate and display individual plots
pp <- plot_scatter("Test_scores", "Mobility", "mediumseagreen", "Test Scores vs Mobility")
p1 <- plot_scatter("Test_scores", "Seg_poverty", "mediumseagreen", "Test Scores vs Poverty")
p2 <- plot_scatter("Test_scores", "Gini", "mediumseagreen", "Test Scores vs Gini")
p3 <- plot_scatter("Test_scores", "Gini_99", "mediumseagreen", "Test Scores vs Gini (99%)")
p4 <- plot_scatter("Test_scores", "Middle_class", "mediumseagreen", "Test Scores vs Middle Class")
p5 <- plot_scatter("Test_scores", "Single_mothers", "mediumseagreen", "Test Scores vs Single Mothers")
p6 <- plot_scatter("Test_scores", "School_spending", "mediumseagreen", "Test Scores vs School Spending")

# Print plots one by one\
print(pp)
```

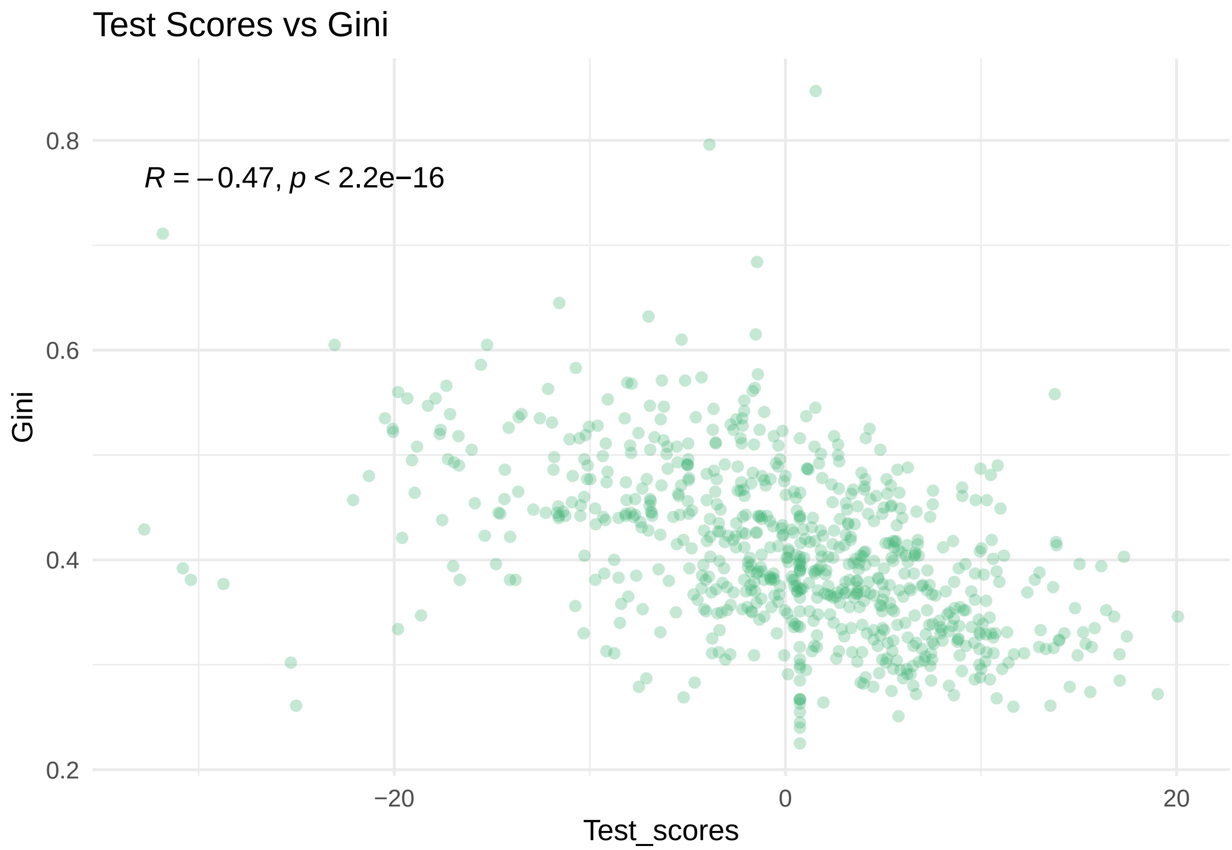

Test Scores vs Mobility
R = 0.41, p < 2.2e−16
0.4
0.3
0.2
0.1
0.0
Mobility
−20
0
20
Test_scores

```
print(p1)
```

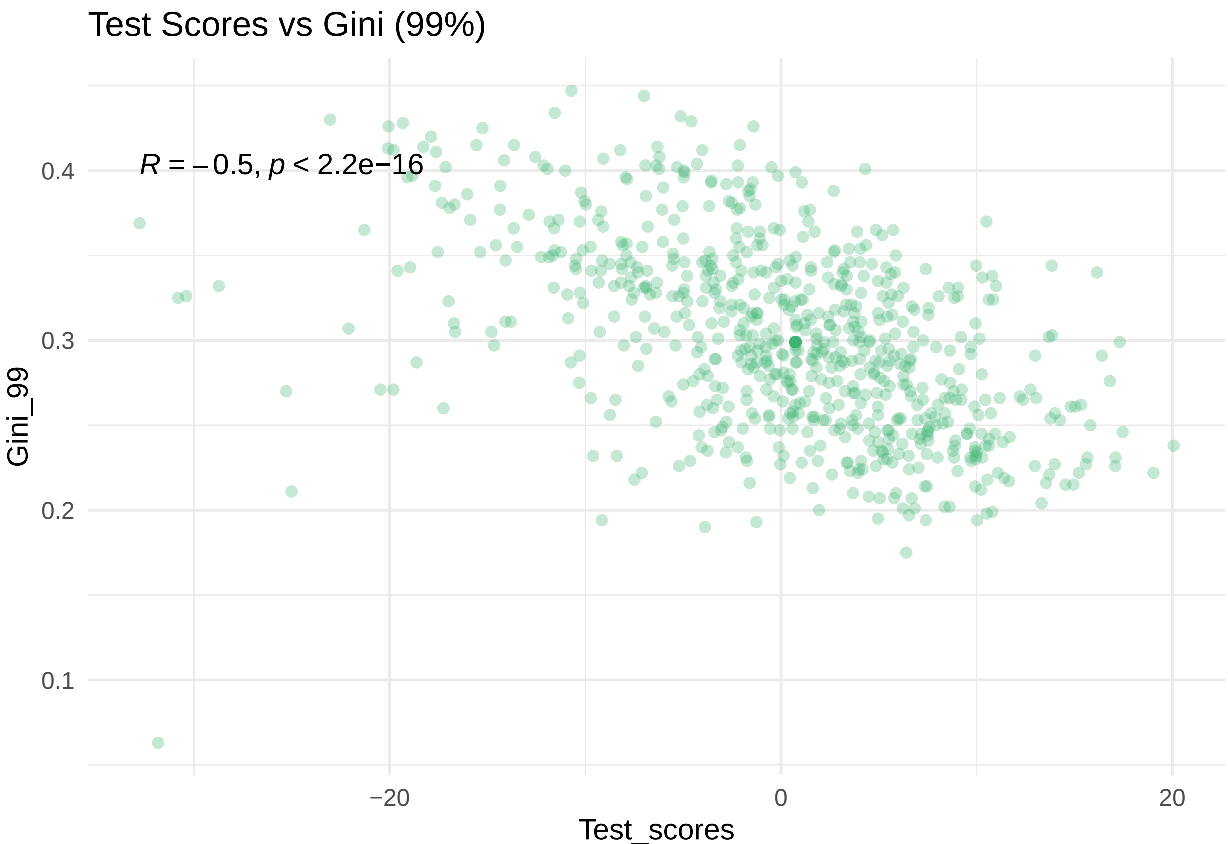


Test Scores vs Poverty
R = −0.25, p = 5.1e−12
0.10
0.05
0.00
Seg_poverty
−20
0
20
Test_scores

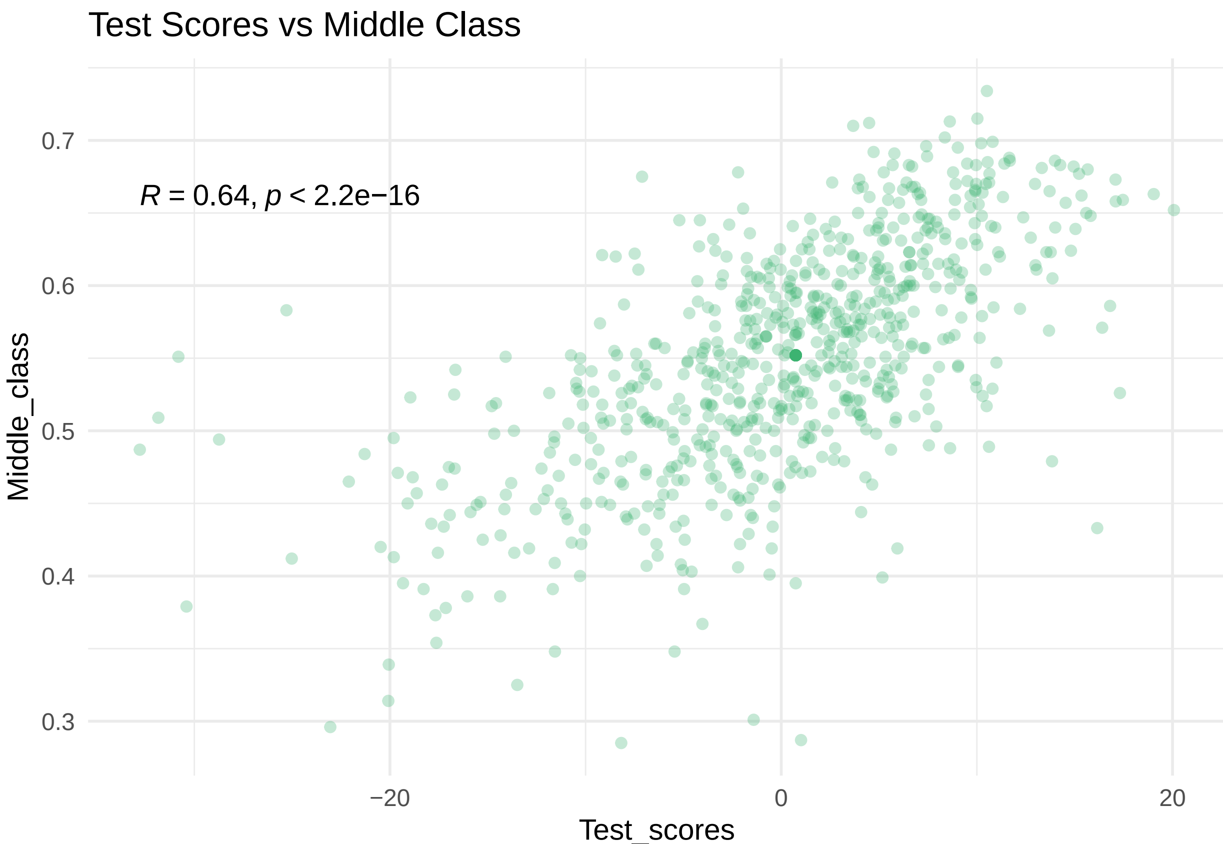

```r
print(p2)
```


Test Scores vs Gini
R = −0.47, p < 2.2e−16
0.8
0.6
0.4
0.2
Gini
−20
0
20
Test_scores

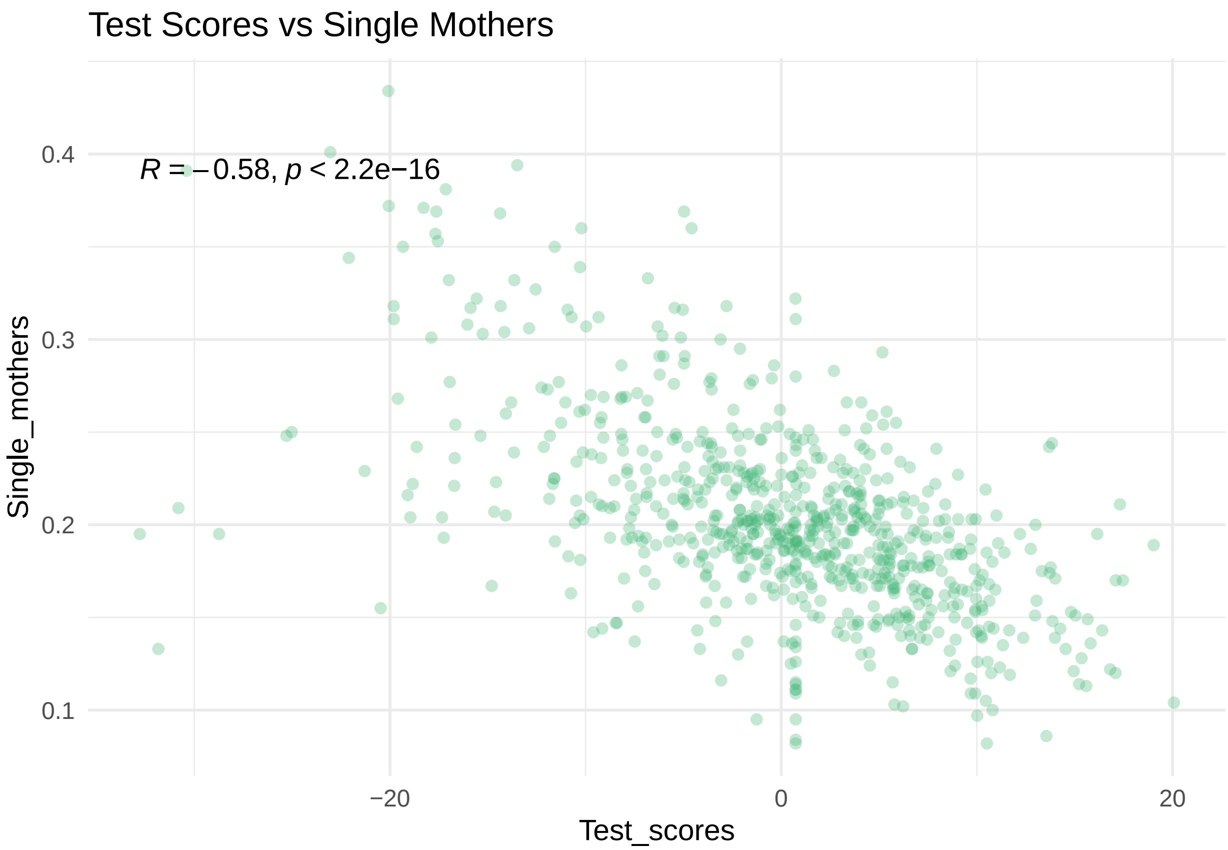
```r
print(p3)
```


Test Scores vs Gini (99%)
R = −0.5, p < 2.2e−16
Gini_99
0.4
0.3
0.2
0.1
−20
0
20
Test_scores

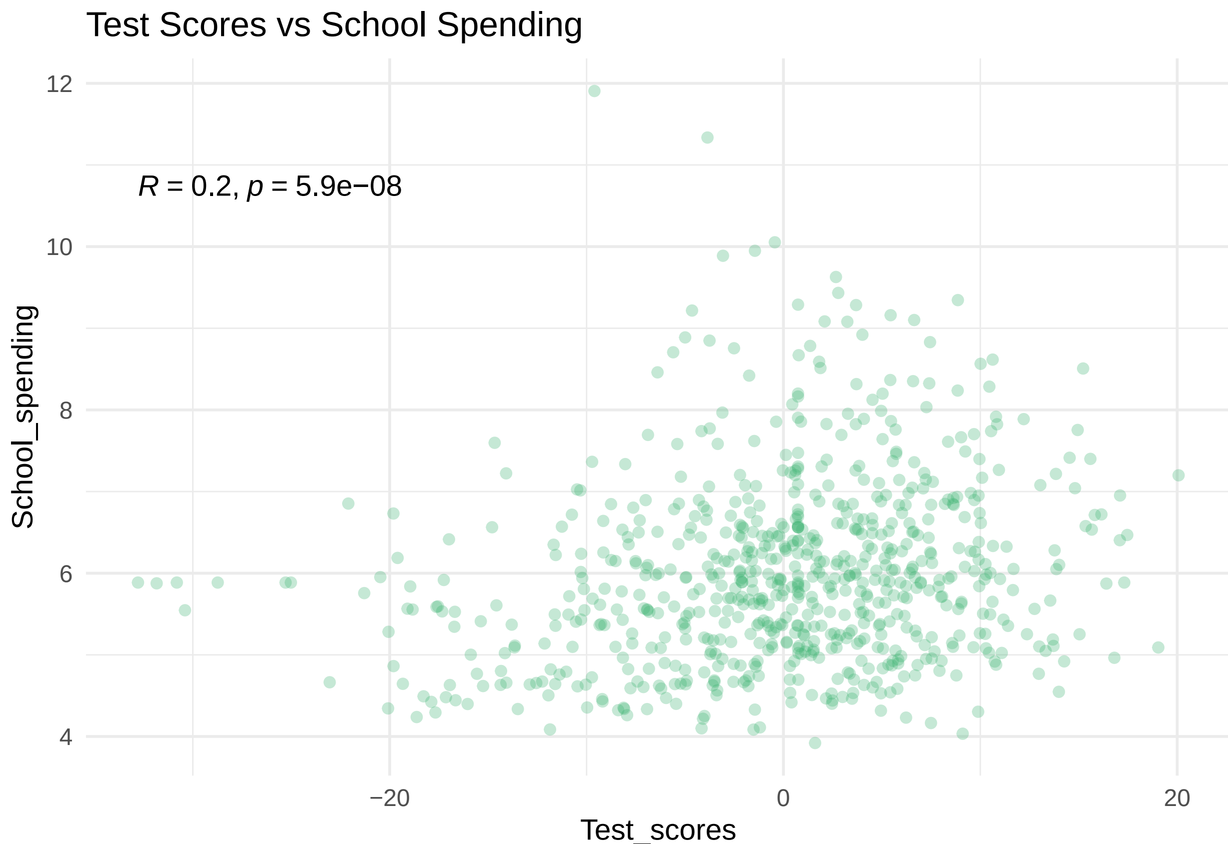
```
print(p4)
```


Test Scores vs Middle Class
R = 0.64, p < 2.2e−16
0.7
0.6
0.5
0.4
0.3
Middle_class
−20
0
20
Test_scores

```r
print(p5)
```


Test Scores vs Single Mothers
R = −0.58, p < 2.2e−16
0.4
0.3
0.2
0.1
Single_mothers
−20
0
20
Test_scores

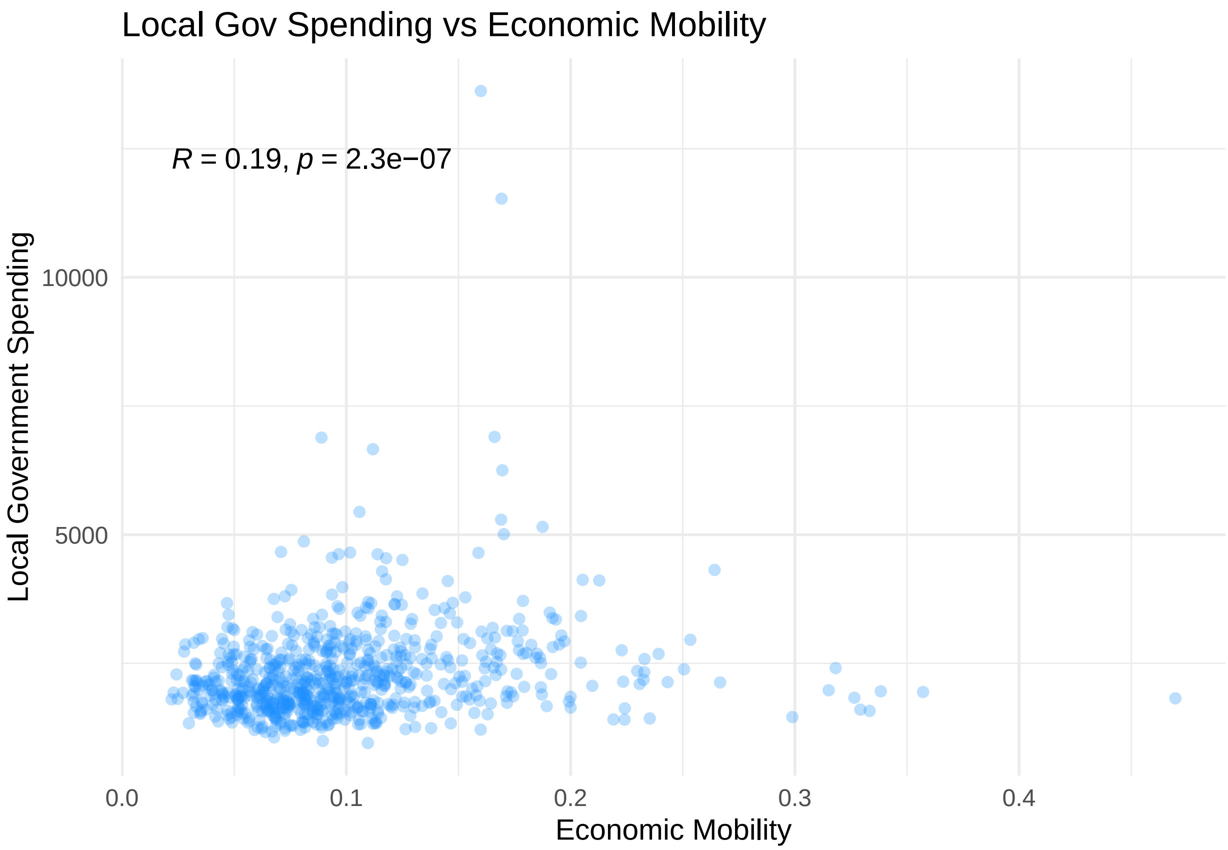
```r
print(p6)
```



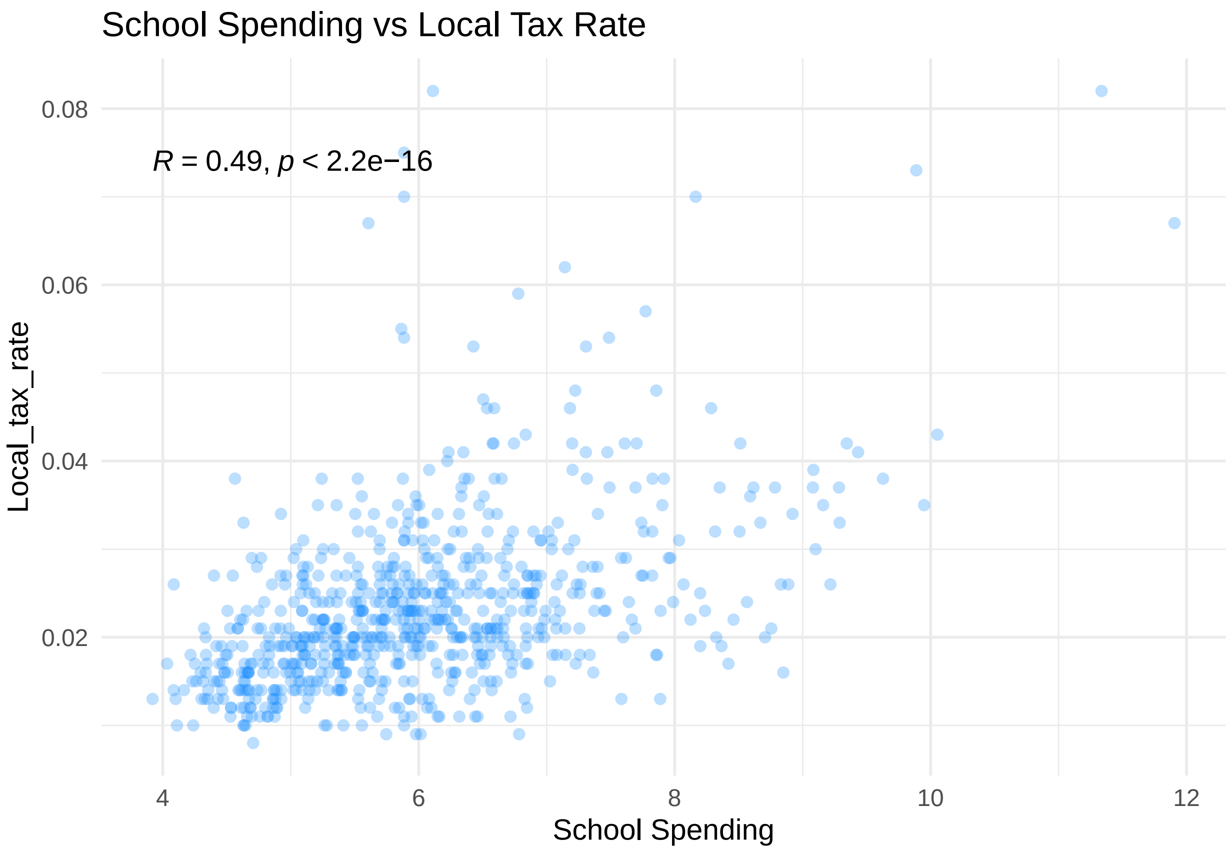

```r
# Define base dataset
data <- cleaned_data  # Use cleaned dataset without missing values

# Plot 1: Mobility vs Local Tax Rate
p1 <- ggplot(data, aes(x = Mobility, y = Local_tax_rate)) +
  geom_point(color = "dodgerblue", alpha = .3) +
  stat_cor(label.x = min(data$Mobility, na.rm = TRUE),
           label.y = max(data$Local_tax_rate, na.rm = TRUE) * 0.9) +
  ggtitle("Local Tax Rate vs Economic Mobility") +
  xlab("Economic Mobility") +
  ylab("Local Tax Rate") +
  theme_minimal()

# Plot 2: Mobility vs Local Government Spending
p2 <- ggplot(data, aes(x = Mobility, y = Local_gov_spending)) +
  geom_point(color = "dodgerblue", alpha = .3) +
  stat_cor(label.x = min(data$Mobility, na.rm = TRUE),
           label.y = max(data$Local_gov_spending, na.rm = TRUE) * 0.9) +
  ggtitle("Local Gov Spending vs Economic Mobility") +
  xlab("Economic Mobility") +
  ylab("Local Government Spending") +
  theme_minimal()

# Print each plot separately
print(p1)
```


Local Tax Rate vs Economic Mobility
R = 0.34, p < 2.2e−16
Local Tax Rate
0.08
0.06
0.04
0.02
0.0
0.1
0.2
0.3
0.4
Economic Mobility

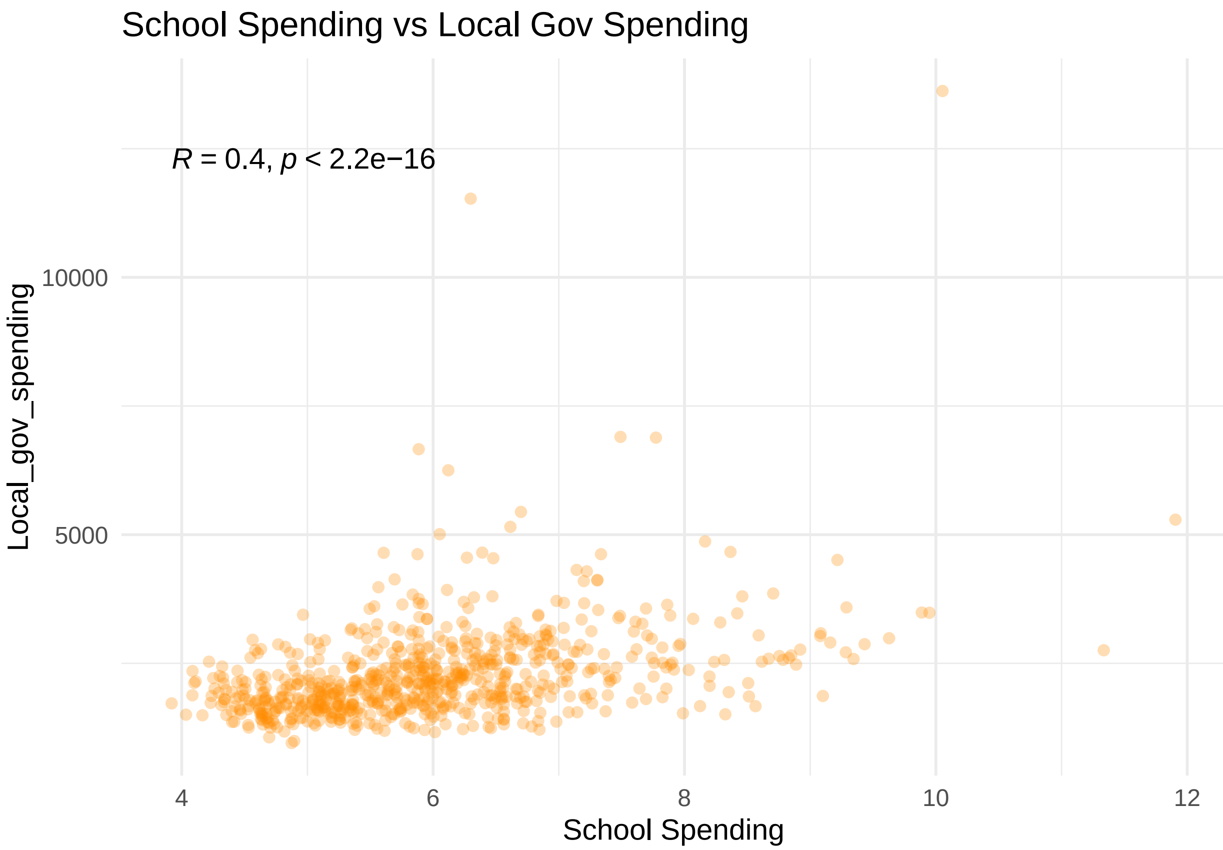
```
print(p2)
```



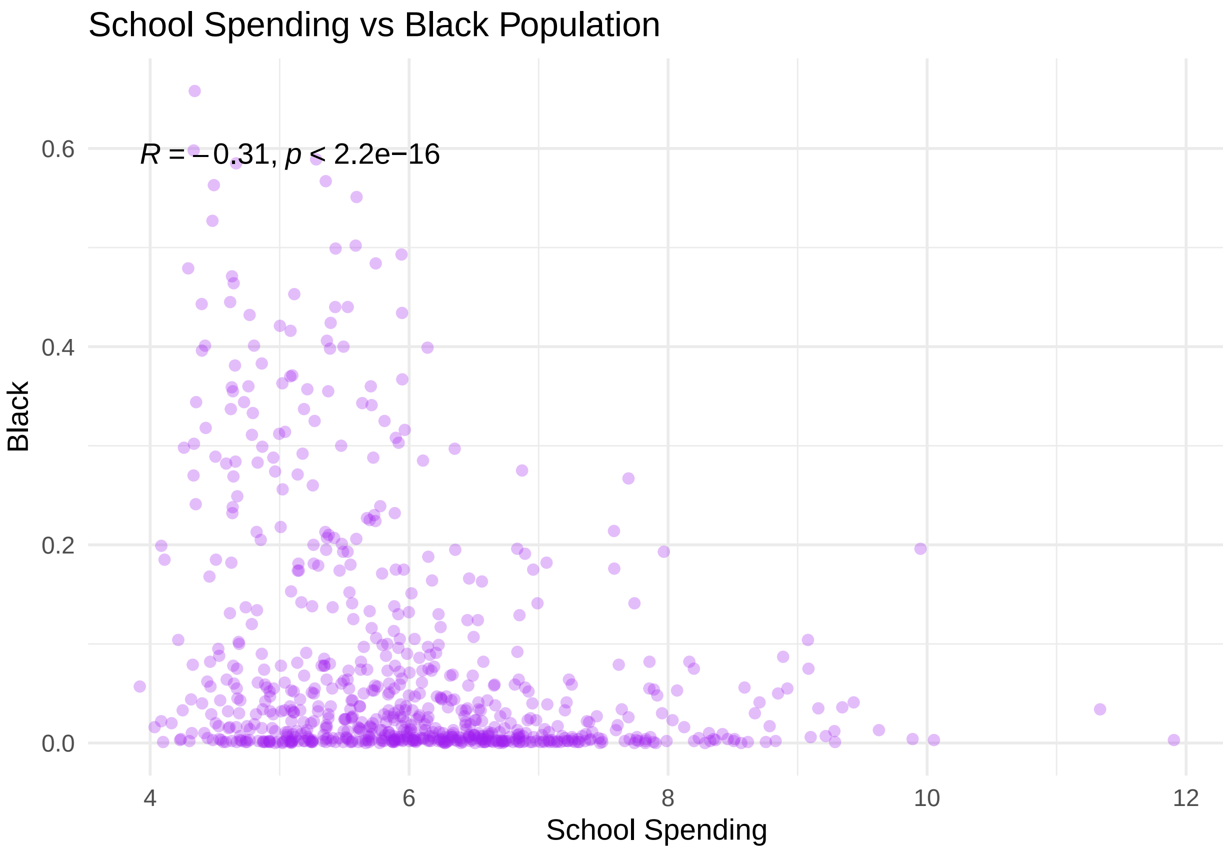

```
# Remove rows with missing or infinite values in relevant columns
data_filtered <- cleaned_data %>%
  filter(
    !is.na(School_spending) & !is.na(Local_tax_rate) & !is.na(Local_gov_spending) & !is.na(Black) &
    is.finite(School_spending) & is.finite(Local_tax_rate) & is.finite(Local_gov_spending) & is.finite(
  )

# Function to create scatter plots
plot_scatter <- function(x_var, color, title) {
  ggplot(data_filtered, aes(x = School_spending, y = .data[[x_var]])) +
    geom_point(color = color, alpha = .3) +
    stat_cor(label.x = min(data_filtered$School_spending, na.rm = TRUE),
             label.y = max(data_filtered[[x_var]], na.rm = TRUE) * 0.9) +
    ggtitle(title) +
    xlab("School Spending") +
    ylab(x_var) +
    theme_minimal()
}

# Generate and display each plot separately
print(plot_scatter("Local_tax_rate", "dodgerblue", "School Spending vs Local Tax Rate"))
```




```r
print(plot_scatter("Local_gov_spending", "darkorange", "School Spending vs Local Gov Spending"))
```




```
print(plot_scatter("Black", "purple", "School Spending vs Black Population"))
```


School Spending vs Black Population
R = −0.31, p < 2.2e−16
Black
0.6
0.4
0.2
0.0
4
6
8
10
12
School Spending