Religion-Voter Turnout Project

Introduction

Read in data

```
library(readxl)
religion <- read_excel("religion_census.xlsx")
head(religion)
```

```
## # A tibble: 6 x 568
##   TOTCNG TOTADH TOTRATE EVANCNG EVANADH EVANRATE BPRTCNG BPRTADH BPRTRATE
##    <dbl>  <dbl>   <dbl>   <dbl>   <dbl>    <dbl>   <dbl>   <dbl>    <dbl>
## 1    106  36938    677.      79   27503     504.      13    2291    42.0
## 2    271  96918    532.     178   57986     318.      17    3130    17.2
## 3     89  15101    550.      51    8793     320.      21    3328   121.
## 4     81  11430    499.      63   10159     443.       8     966    42.2
## 5    156  37352    652.     122   29223     510.       2      58     1.01
## 6     47   6300    577.      15    2152     197.      21    3445   316.
## # ... with 559 more variables: MPRTCNG <dbl>, MPRTADH <dbl>, MPRTRATE <dbl>,
## #   CATHCNG <dbl>, CATHADH <dbl>, CATHRATE <dbl>, ORTHCNG <dbl>, ORTHADH <dbl>,
## #   ORTHRATE <dbl>, OTHCNG <dbl>, OTHADH <dbl>, OTHRATE <dbl>, OCGCNG <dbl>,
## #   AMECNG <dbl>, AMEADH <dbl>, AMERATE <dbl>, AMEZCNG <dbl>, AMEZADH <dbl>,
## #   AMEZRATE <dbl>, ALBCNG <dbl>, ALBADH <dbl>, ALBRATE <dbl>, AWMCCNG <dbl>,
## #   AWMCADH <dbl>, AWMCRATE <dbl>, ALBPCNG <dbl>, AMANCNG <dbl>, AMANADH <dbl>,
## #   AMANRATE <dbl>, AAMCNG <dbl>, AAMADH <dbl>, AAMRATE <dbl>, ...
```

We only want data on predominantly African-American religious institutions to analyze the effects of Black Churches on voter turnout in the South. And we will narrow the data to a few states in the deep south: South Carolina, Georgia, Louisiana, Alabama, and Mississippi.

To understand what churches have the greatest impact, the Conference of National Black Churches provides a starting point. The organization said it represents more than 80% of African-American Christians, or around 20 million people in the U.S. It is made up of seven predominantly African-American Churches: - African Methodist Episcopal Church (AME) - African Methodist Episcopal Zion Church (AMEZ) - Christian Methodist Episcopal Church (CME) - Church of God In Christ (COGIC) - National Baptist Convention of America, Inc., International (NBCA) - National Baptist Convention, U. S. A. Inc. (NBC USA) - Progressive National Baptist Convention, Inc. (PNBC)

However, the convention does not include Black Protestants, which we will add on to the dataset. We now go into the dataset to filter for these denominations.

Filtering data

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
```

```
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
#identify the abbreviations of 5 states we want to include
southern_states <- c("SC", "GA", "LA", "AL", "MS")
#identified the denominations' codes in the codebook
aa_denominations <- c("AMECNG", "AMEADH", "AMERATE", "AMEZCNG", "AMEZADH", "AMEZRATE", "CMECNG", "CMEADH
other_variables <- c("TOTCNG", "TOTADH", "TOTRATE", "FIPS", "STABBR", "CNTYNAME", "CNTYCODE", "POP2010")

southern_aa <- religion[c(other_variables, aa_denominations)] %>%
  filter(STABBR  %in% southern_states) #data frame with all the African-American churches in the 5 stat
```

```
#we check if the number of counties in the dataframe matches up with the number of counties there are i
#LA has 64 states, AL has 67, GA has 159, SC has 46 and MS has 82.
nrow(southern_aa) == 64 + 67 + 159 + 46 + 82
```

Sanity check

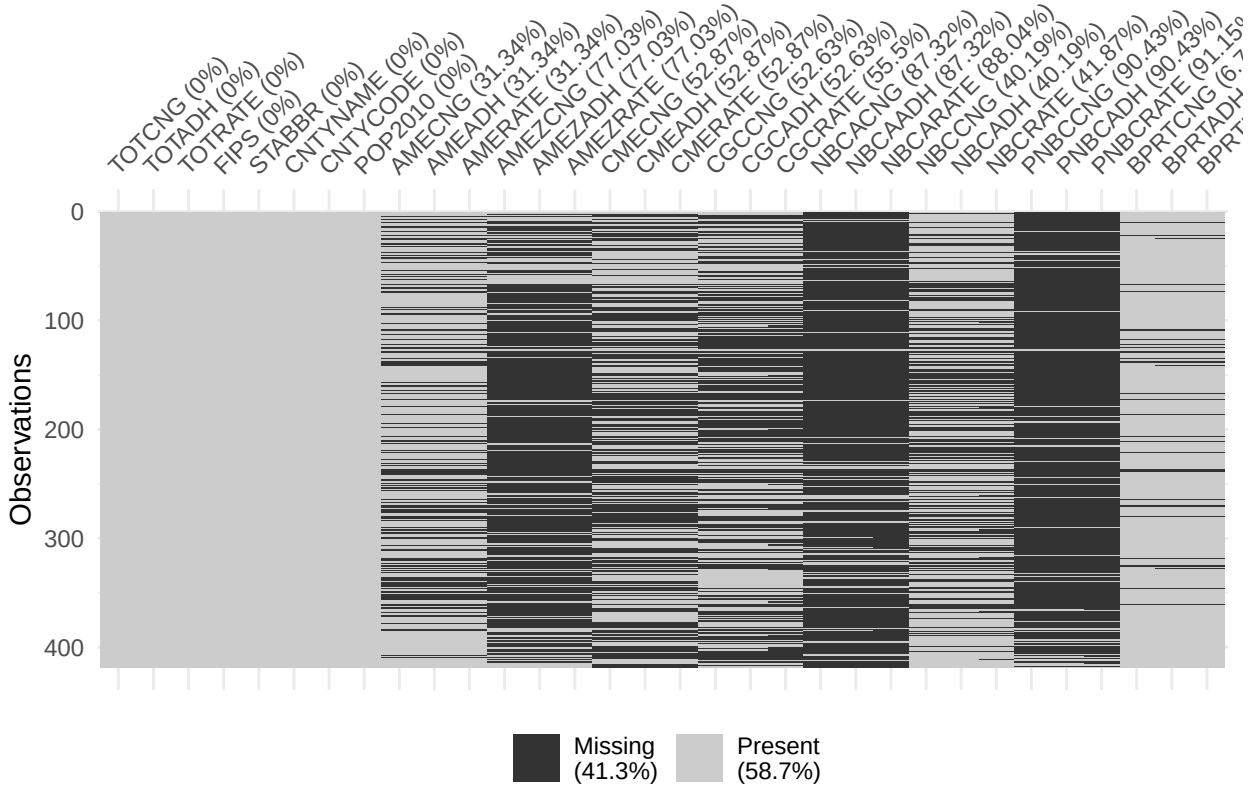
```
## [1] TRUE
```

```
#we also check whether total adherents are less than total adherents in the county
adh_var <- c("AMEADH", "AMEZADH", "CMEADH", "CGCADH", "NBCAADH", "NBCADH", "PNBCADH", "BPRTADH")
for (i in adh_var) {
  if (any(southern_aa[[i]] > southern_aa$TOTADH, na.rm = TRUE)) {i}
}
#check if total adherents are less than total population
which(southern_aa$TOTADH > southern_aa$POP2010)
```

```
## [1] 371
```

Clean up NAs

```
library(visdat)
vis_miss(southern_aa)
```



We observe that the dataset has a lot of missing data, attributed to the fact that some counties simply do not have an predominantly African-American church present. *Note: There are a few denominations that has a higher missing data rate for their rate of adherents than they do for the total congregations variable. Namely: CGCC (Church of God in Christ), NBCA (National Baptist Convention of America Inc.) NBC (National Baptist Convention U.S.A), PNBC (Progressive National Baptist Convention, Inc.), BPRT (Black Protestants)* We now examine what accounts for those inconsistencies.

```
cng_var_miss <- c("CGCCNG", "NBCACNG", "NBCCNG", "PNBCCNG", "BPRTCNG")
adh_var_miss <- c("CGCADH", "NBCAADH", "NBCADH", "PNBCADH", "BPRTADH")
rate_var_miss <- c("CGCRATE", "NBCARATE", "NBCRATE", "PNBCRATE", "BPRTRATE")
check_rate <- vector(length = 3)
for (i in 1:5){
  print(southern_aa[!is.na(southern_aa[cng_var_miss[i]]) & is.na(southern_aa[rate_var_miss[i]]), c(cng_
}
```

```
## # A tibble: 12 x 3
##    CGCCNG CGCADH CGCRATE
##     <dbl>  <dbl>   <dbl>
##  1      1      0      NA
##  2      1      0      NA
##  3      1      0      NA
##  4      1      0      NA
##  5      1      0      NA
##  6      1      0      NA
##  7      1      0      NA
##  8      1      0      NA
##  9      1      0      NA
## 10      1      0      NA
## 11      1      0      NA
## 12      1      0      NA
```

```
## # A tibble: 3 x 3
##   NBCACNG NBCAADH NBCARATE
##     <dbl>   <dbl>    <dbl>
## 1       1       0       NA
## 2       1       0       NA
## 3       1       0       NA
## # A tibble: 7 x 3
##   NBCCNG NBCADH NBCRATE
##    <dbl>  <dbl>   <dbl>
## 1      1      0      NA
## 2      1      0      NA
## 3      1      0      NA
## 4      1      0      NA
## 5      1      0      NA
## 6      1      0      NA
## 7      1      0      NA
## # A tibble: 3 x 3
##   PNBCCNG PNBCADH PNBCRATE
##     <dbl>   <dbl>    <dbl>
## 1       1       0       NA
## 2       1       0       NA
## 3       1       0       NA
## # A tibble: 3 x 3
##   BPRTCNG BPRTADH BPRTRATE
##     <dbl>   <dbl>    <dbl>
## 1       1      NA       NA
## 2       1      NA       NA
## 3       1      NA       NA
```

```
#the adherents numbers in these counties' denominations are 0, therefore we can code the rate as 0.
```

```
cng_var <- c("AMECNG", "AMEZCNG", "CMECNG", "CGCCNG", "NBCACNG", "NBCCNG", "PNBCCNG", "BPRTCNG")
table(apply(is.na(southern_aa[cng_var]), 1, sum))
```

```
##
##   0   1   2   3   4   5   6   7   8
##   7   9  38  66 103  95  65   7  28
```

The table examines how many denominations are absent in any given county. We find that 7 counties have all denominations present and 8 counties has none.

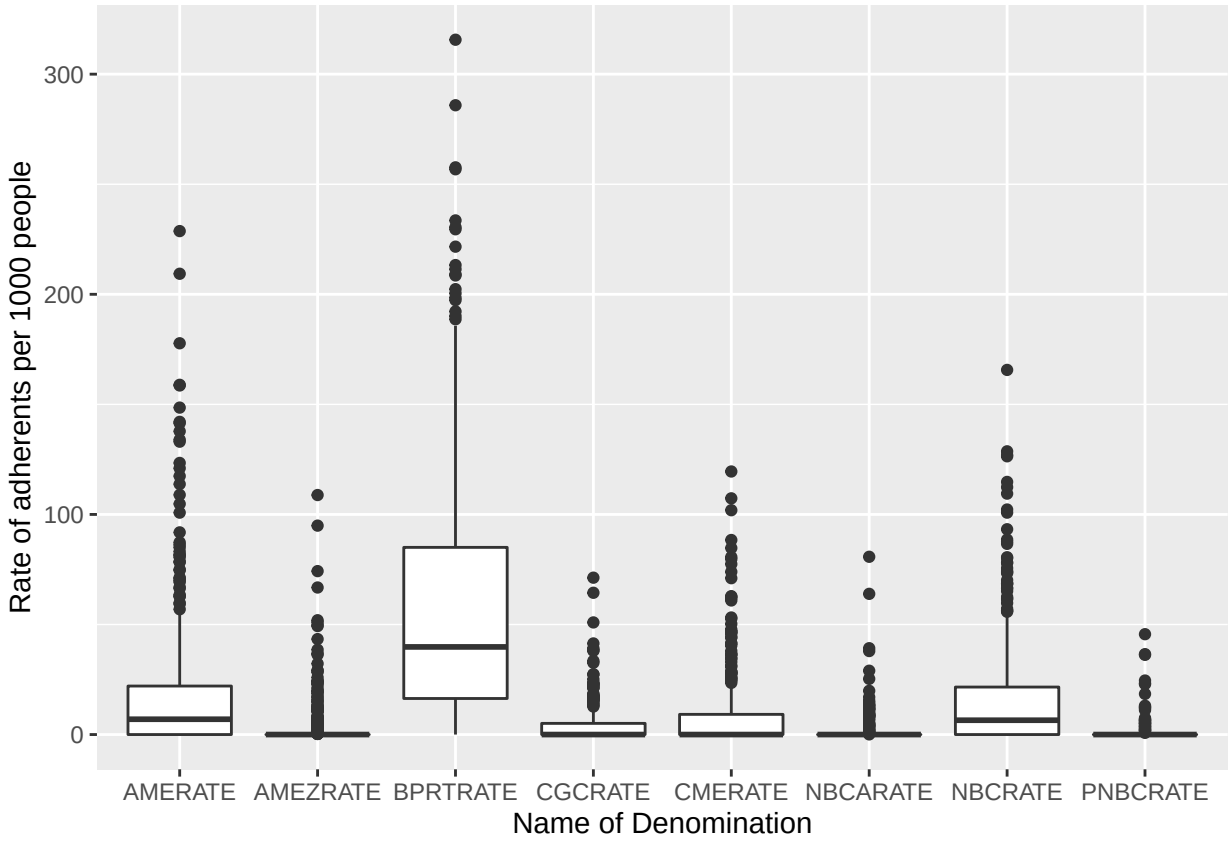
```
southern_aa[is.na(southern_aa)] <- 0
#verify that the dataset now has no missing values
```

We want to combine all the data on the denominations into one coherent set of data for the number of predominently African-American congregations, adherents and adherent rates for each county. This makes it easier to conduct analysis but also could expose the data to variations across different denominations.

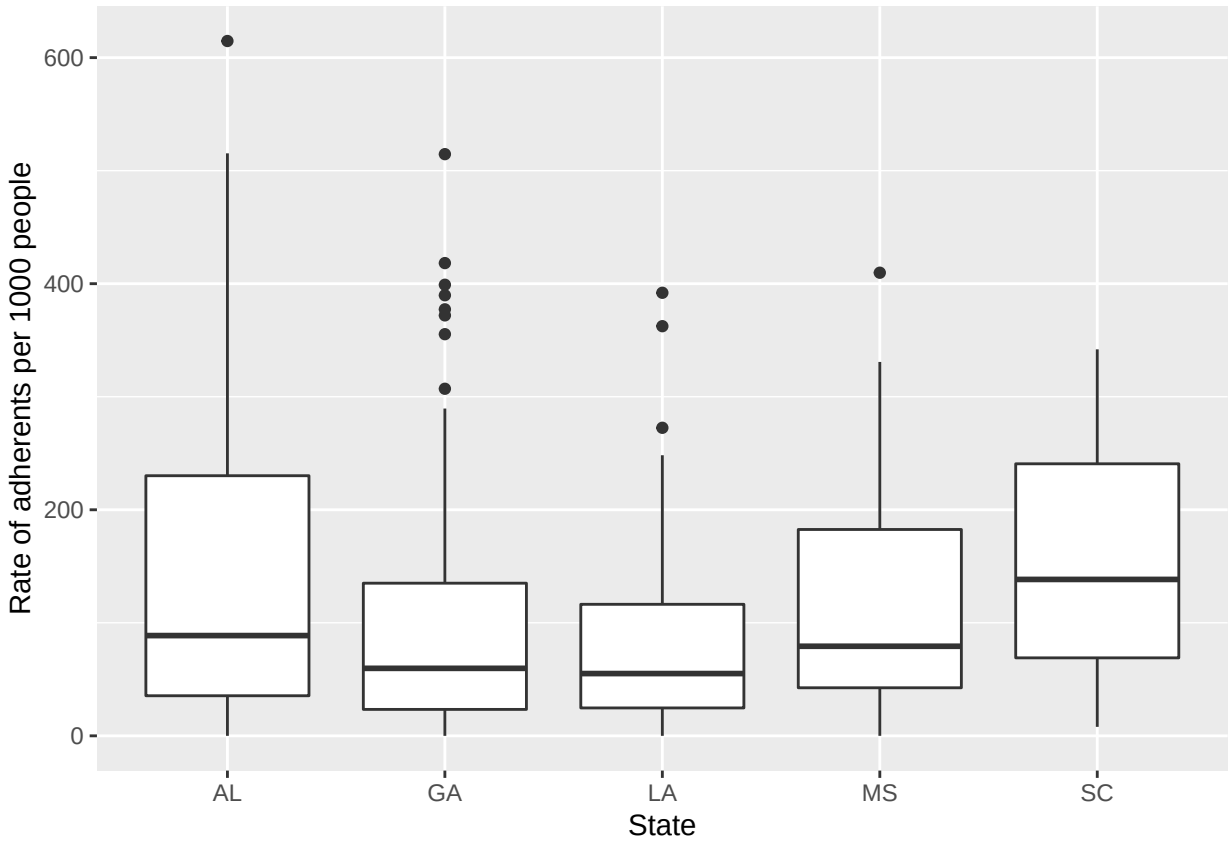
```
rate_var <- c("AMERATE", "AMEZRATE", "CMERATE", "CGCRATE", "NBCARATE", "NBCRATE", "PNBCRATE", "BPRTRATE
cng_var <- c("AMECNG", "AMEZCNG", "CMECNG", "CGCCNG", "NBCACNG", "NBCCNG", "PNBCCNG", "BPRTCNG")
southern_aa$AA_CNG <- apply(southern_aa[cng_var], 1, sum)
southern_aa$AA_ADH <- apply(southern_aa[adh_var], 1, sum)
southern_aa$AA_RATE <- apply(southern_aa[rate_var], 1, sum)
```

EDA on the Religion dataset

```
library(ggplot2)
#boxplots of denominations' adherents rate in counties
rate_var <- c("AMERATE", "AMEZRATE", "CMERATE", "CGCRATE", "NBCARATE", "NBCRATE", "PNBCRATE", "BPRTRATE
rate_data <- southern_aa %>%
  gather(key = "denomination", value = "rate", c(11, 14, 17, 20, 23, 26, 29, 32))
ggplot(data = rate_data, aes(x = denomination, y = rate)) +
  geom_boxplot() +
  labs(x = "Name of Denomination", y = "Rate of adherents per 1000 people")
```

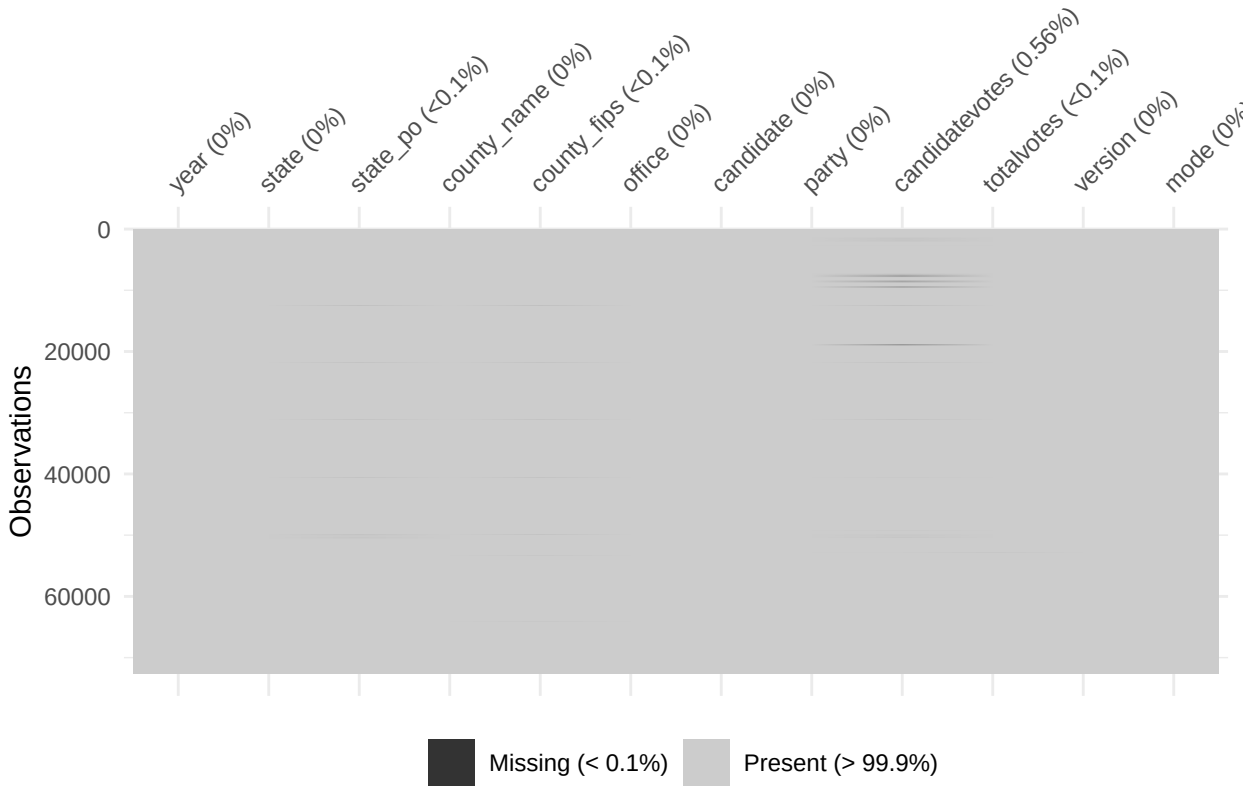



```
ggplot(data = southern_aa, aes(x = STABBR, y = AA_RATE)) +
  geom_boxplot() +
  labs(x = "State", y = "Rate of adherents per 1000 people")
```



Import election result data by county

```
library(visdat)
pres_county <- read.csv("countypres_2000-2020.csv")
vis_miss(pres_county)
```




```
which(is.na(pres_county$county_fips)) %in% southern_aa$FIPS
```

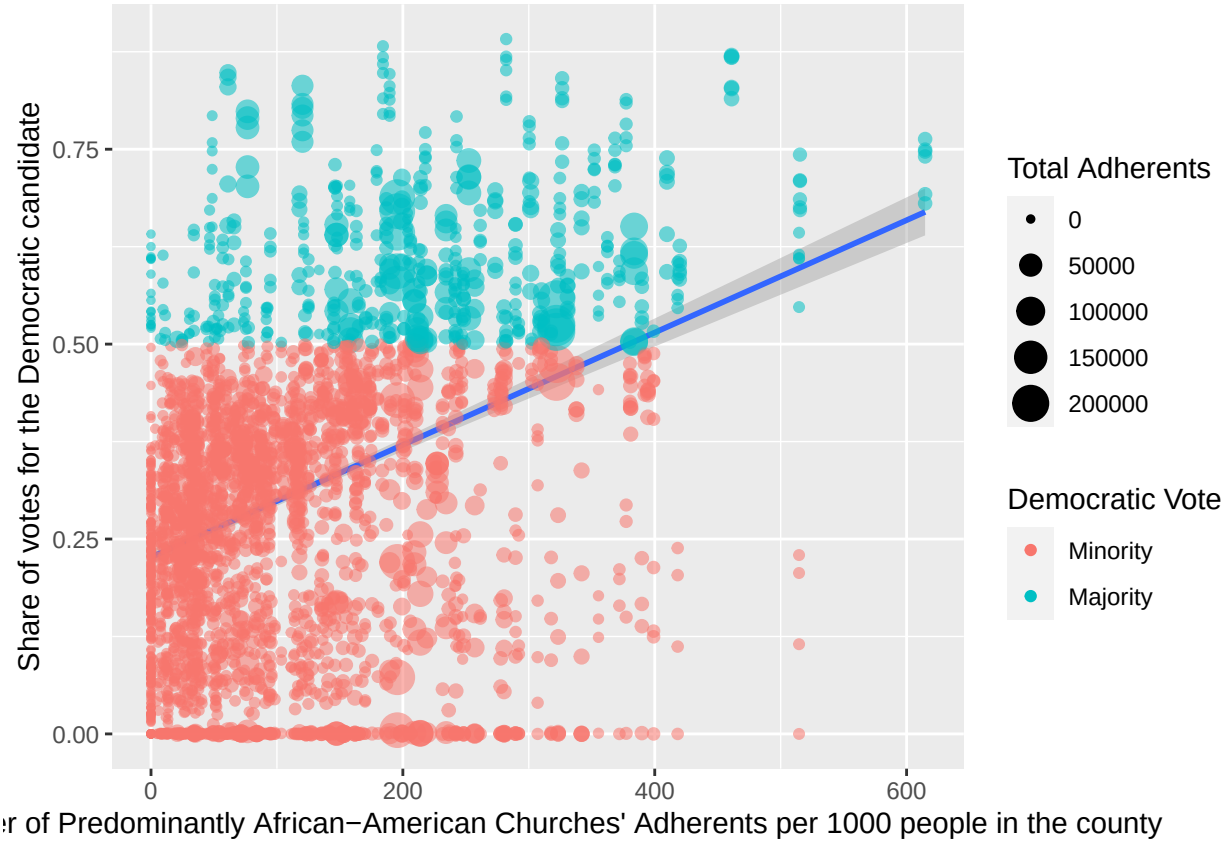
```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
#none of the missing fips are in the southern states dataset
pres_southern <- right_join(pres_county, southern_aa, by = c("county_fips" = "FIPS"))
pres_clean <- pres_southern %>%
  filter(party == "DEMOCRAT") %>%
  mutate(dem_share = candidatevotes/totalvotes) %>%
  select(-c("office", "candidate", "party", "version", "mode", "county_name", "state_po", "state"))
```

Visualize

```
outcome <- pres_clean$dem_share > 0.5
outcome_char <- as.character(outcome)
outcome_char[outcome] <- "Majority"
outcome_char[!outcome] <- "Minority"
outcome_char <- factor(outcome_char, levels = c("Minority", "Majority"))
g_pres_church <- ggplot(data = pres_clean, aes(x = AA_RATE, y = dem_share))
g_pres_church +
  geom_smooth(method = "lm") +
  geom_point(aes(alpha = 0.05, color = outcome_char, size = AA_ADH)) +
  guides(alpha = "none") +
  labs(x = "Number of Predominantly African-American Churches' Adherents per 1000 people in the county"
```

```
## `geom_smooth()` using formula 'y ~ x'
```



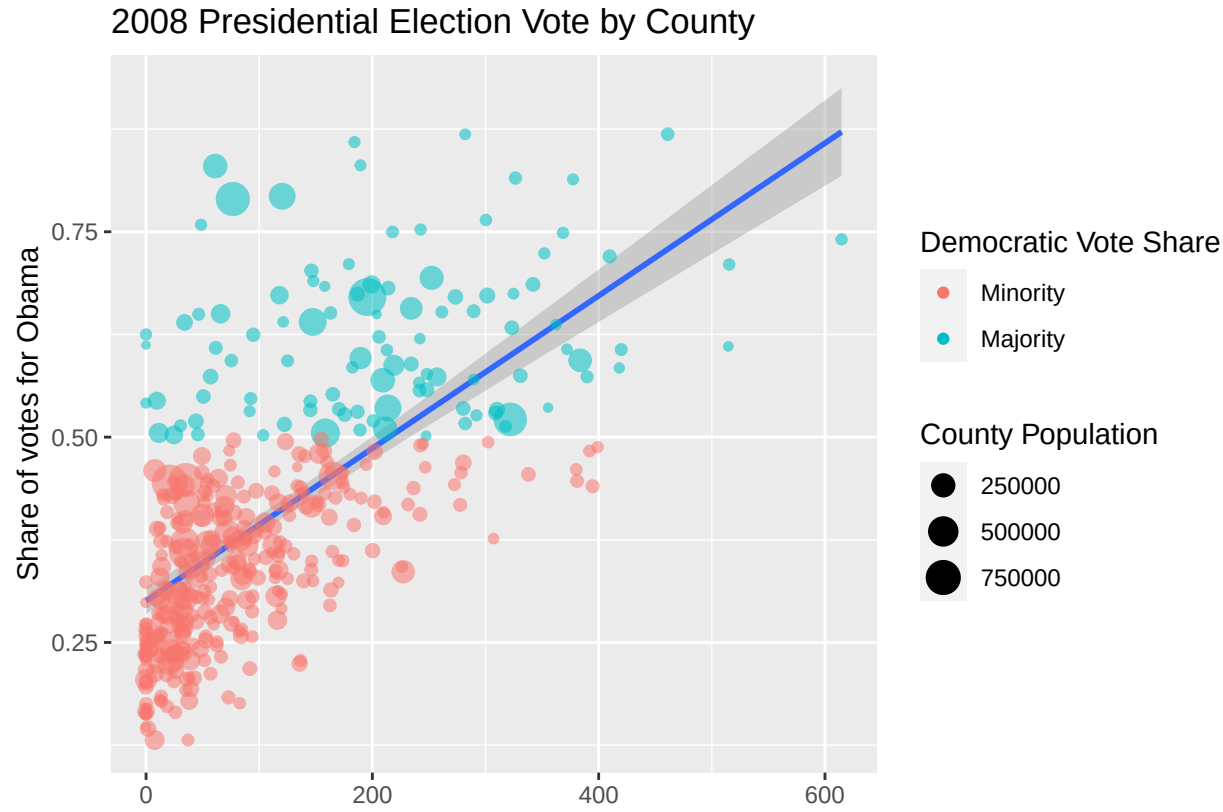
#2008 & 2012 performance of Presidential candidates

```
#2008 vote
pres_2008 <- pres_clean %>%
  filter(year == 2008)

#encode the outcome variable as dem majority/minority
outcome_2008 <- pres_2008$dem_share > 0.5
outcome_2008_char <- as.character(outcome_2008)
outcome_2008_char[outcome_2008] <- "Majority"
outcome_2008_char[!outcome_2008] <- "Minority"
outcome_2008_char <- factor(outcome_2008_char, levels = c("Minority", "Majority"))

#graph rate vs. dem_share
g_pres_church_2008 <- ggplot(data = pres_2008, aes(x = AA_RATE, y = dem_share))
g_pres_church_2008 +
  geom_smooth(method = "lm") +
  geom_point(aes(alpha = 0.05, color = outcome_2008_char, size = POP2010)) +
  guides(alpha = "none") +
  labs(x = "Number of Predominantly African-American Churches' Adherents per 1000 people in the county"
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggsave("scatterplot_2018.png", width = 10, height = 6)
```

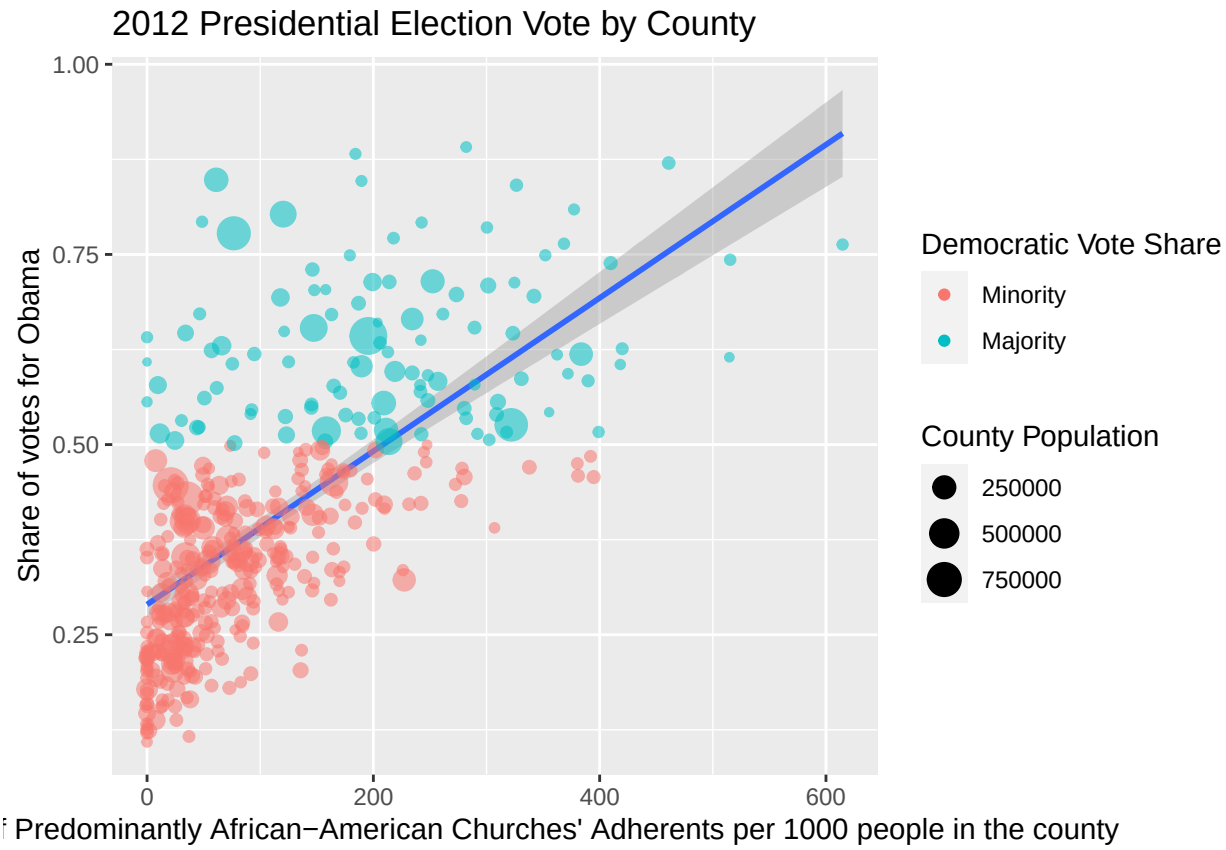
```
## `geom_smooth()` using formula 'y ~ x'
```

```
#2012 vote
pres_2012 <- pres_clean %>%
  filter(year == 2012)

#encode the outcome variable as dem majority/minority
outcome_2012 <- pres_2012$dem_share > 0.5
outcome_2012_char <- as.character(outcome_2012)
outcome_2012_char[outcome_2012] <- "Majority"
outcome_2012_char[!outcome_2012] <- "Minority"
outcome_2012_char <- factor(outcome_2012_char, levels = c("Minority", "Majority"))

#graph rate vs. dem_share
g_pres_church_2012 <- ggplot(data = pres_2012, aes(x = AA_RATE, y = dem_share))
g_pres_church_visa <- g_pres_church_2012 +
  geom_smooth(method = "lm") +
  geom_point(aes(alpha = 0.05, color = outcome_2012_char, size = POP2010)) +
  guides(alpha = "none") +
  labs(x = "Number of Predominantly African-American Churches' Adherents per 1000 people in the county"
g_pres_church_visa
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
#ggsave("scatterplot.png", width = 10, height = 6)
```

```
fit <- lm(pres_2012$dem_share ~ pres_2012$AA_RATE)
summary(fit)
```

Try fitting the model

```
##
## Call:
## lm(formula = pres_2012$dem_share ~ pres_2012$AA_RATE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23015 -0.08681 -0.02478  0.06803  0.49681
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.897e-01  8.865e-03   32.68   <2e-16 ***
## pres_2012$AA_RATE 1.008e-03  5.655e-05   17.82   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1255 on 416 degrees of freedom
## Multiple R-squared:  0.4329, Adjusted R-squared:  0.4316
## F-statistic: 317.6 on 1 and 416 DF,  p-value: < 2.2e-16
```

```r
fit2 <- lm(pres_2008$dem_share ~ pres_2008$AA_RATE)
summary(fit2)
```

```
##
## Call:
## lm(formula = pres_2008$dem_share ~ pres_2008$AA_RATE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.22698 -0.08357 -0.02392  0.06019  0.47269
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.004e-01  8.302e-03   36.19   <2e-16 ***
## pres_2008$AA_RATE 9.292e-04  5.295e-05   17.55   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1175 on 416 degrees of freedom
## Multiple R-squared:  0.4253, Adjusted R-squared:  0.424
## F-statistic: 307.9 on 1 and 416 DF,  p-value: < 2.2e-16
```

Case Study: Louisiana

```r
Louisiana_2012_raw <- read_excel("LA_2012_1106_sta.xls", sheet = 1)
```

```
## New names:
## * `` -> ...2
## * `` -> ...3
## * `` -> ...4
## * `` -> ...5
## * `` -> ...6
## * ...
```

```r
Louisiana_2012_raw <- Louisiana_2012_raw %>%
  slice(-c(1:6, 9)) %>%
  select(-c(1,3,5,7,9,11,13,15,17,19))

colnames(Louisiana_2012_raw) <- c("Jurisdiction", "Total", "Black_total_reg", "Dem_total_reg", "Black_d

Louisiana_2012_raw1 <- Louisiana_2012_raw %>% slice(-c(1,2))

Louisiana_2012_voting_total <- Louisiana_2012_raw1 %>%
  filter(Jurisdiction == "# VOTING") %>%
  select(-1) %>%
  apply(MARGIN = 2, FUN = as.integer) %>%
  as.data.frame()
colnames(Louisiana_2012_voting_total) <- c("Total_vote", "Black_total_vote", "Dem_total_vote", "Black_d

Louisiana_2012_vote_prop <- Louisiana_2012_raw1 %>%
  filter(Jurisdiction == "% VOTING") %>%
  select(-1) %>%
  apply(MARGIN = 2, FUN = as.integer) %>%
  as.data.frame()
```

```
colnames(Louisiana_2012_vote_prop) <- c("Total_vote_prop", "Black_total_vote_prop", "Dem_total_vote_prop

Louisiana_2012_registration <- Louisiana_2012_raw1 %>%
  filter(Jurisdiction != "# VOTING" & Jurisdiction != "% VOTING")

Louisiana_2012_general <- cbind(Louisiana_2012_registration, Louisiana_2012_voting_total, Louisiana_2012
Louisiana_2012_general$Total <- as.numeric(Louisiana_2012_general$Total)
```

Plotting Louisiana dem_share against rate of predonimantly AA churches adherents

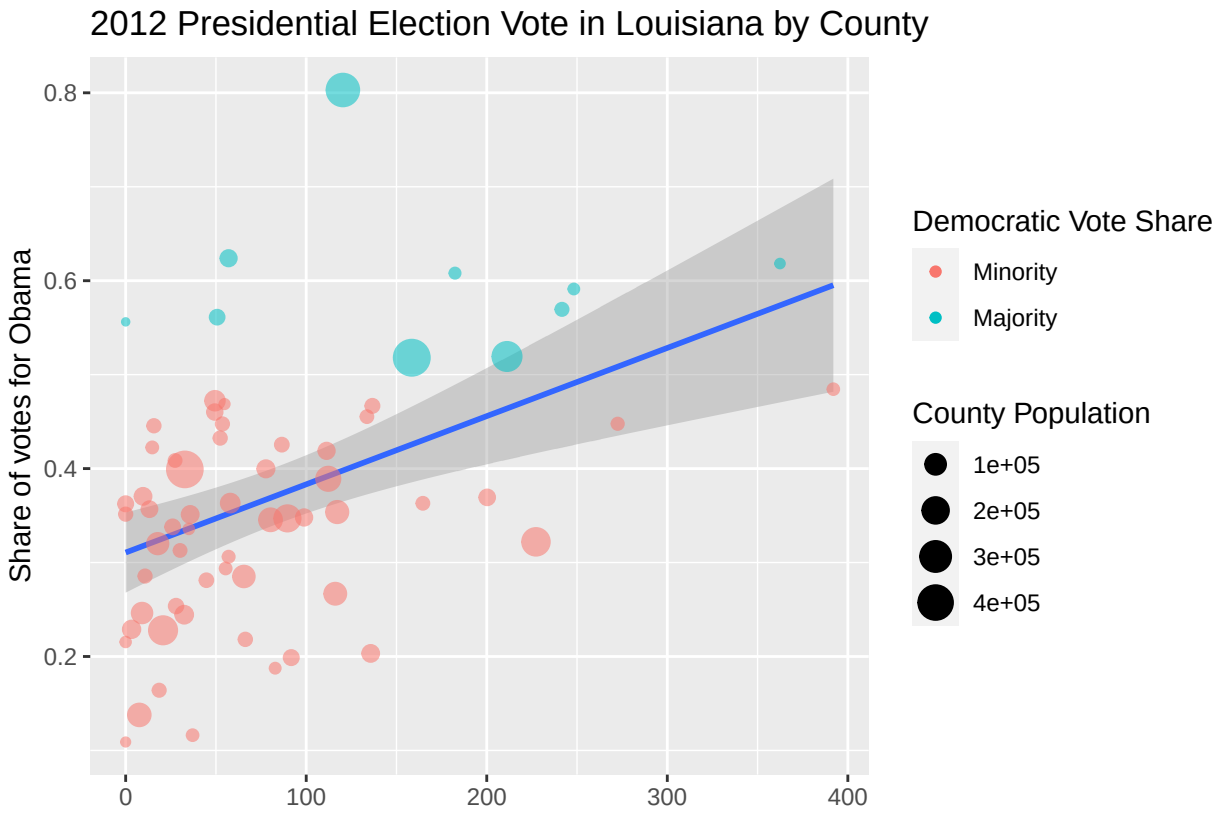
```
pres_LA_2012 <- pres_clean %>%
  filter(STABBR == "LA" & year == 2012) %>%
  select(-c("CNTYNAME", "year")) %>%
  mutate(Jurisdiction = Louisiana_2012_general[-1,]$Jurisdiction)

Louisiana_general_AA <- right_join(Louisiana_2012_general[-1,], pres_LA_2012, by = "Jurisdiction")

#graph rate vs. dem_share
outcome_LA_2012 <- Louisiana_general_AA$dem_share > 0.5
outcome_LA_2012_char <- as.character(outcome_LA_2012)
outcome_LA_2012_char[outcome_LA_2012] <- "Majority"
outcome_LA_2012_char[!outcome_LA_2012] <- "Minority"
outcome_LA_2012_char <- factor(outcome_LA_2012_char, levels = c("Minority", "Majority"))

g_pres_LA_2012 <- ggplot(data = Louisiana_general_AA, aes(x = AA_RATE, y = dem_share))
g_pres_LA_2012 +
  geom_smooth(method = "lm") +
  geom_point(aes(alpha = 0.05, color = outcome_LA_2012_char, size = POP2010)) +
  guides(alpha = "none") +
  labs(x = "Number of Predominantly African-American Churches' Adherents per 1000 people in the county"
```

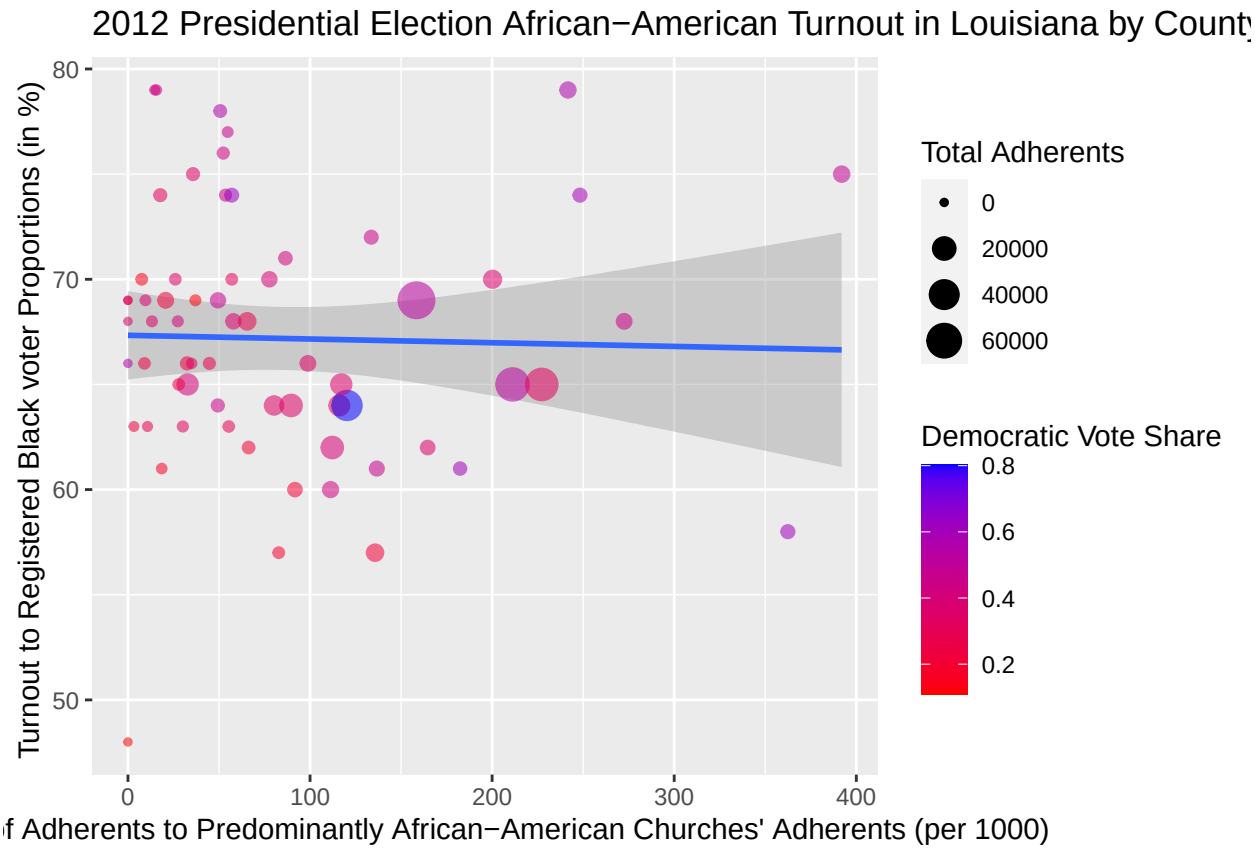
```
## `geom_smooth()` using formula 'y ~ x'
```



Plot Turnout against Rate of Adherents to Predominantly African-American Churches

```
g_turnout_LA_2012 <- ggplot(data = Louisiana_general_AA, aes(x = AA_RATE, y = Black_total_vote_prop))
g_turnout_LA_2012 +
  geom_smooth(method = "lm") +
  geom_point(aes(alpha = 0.05, color = dem_share, size = AA_ADH)) +
  guides(alpha = "none") +
  labs(x = "Rate of Adherents to Predominantly African-American Churches' Adherents (per 1000)", y = "Tu
  scale_color_gradient(low = "red", high = "blue")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

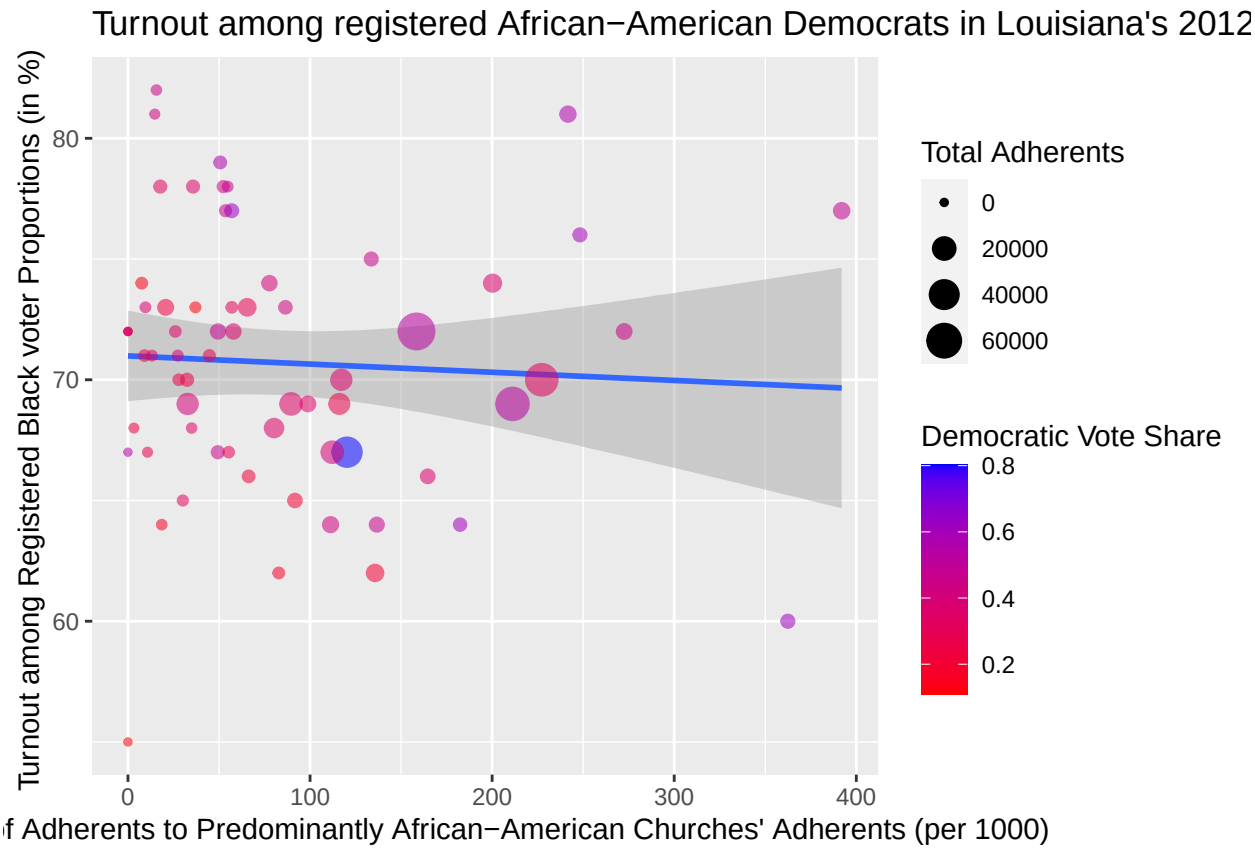



```
ggsave("louisiana_turnout.png", width = 10, height = 6)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

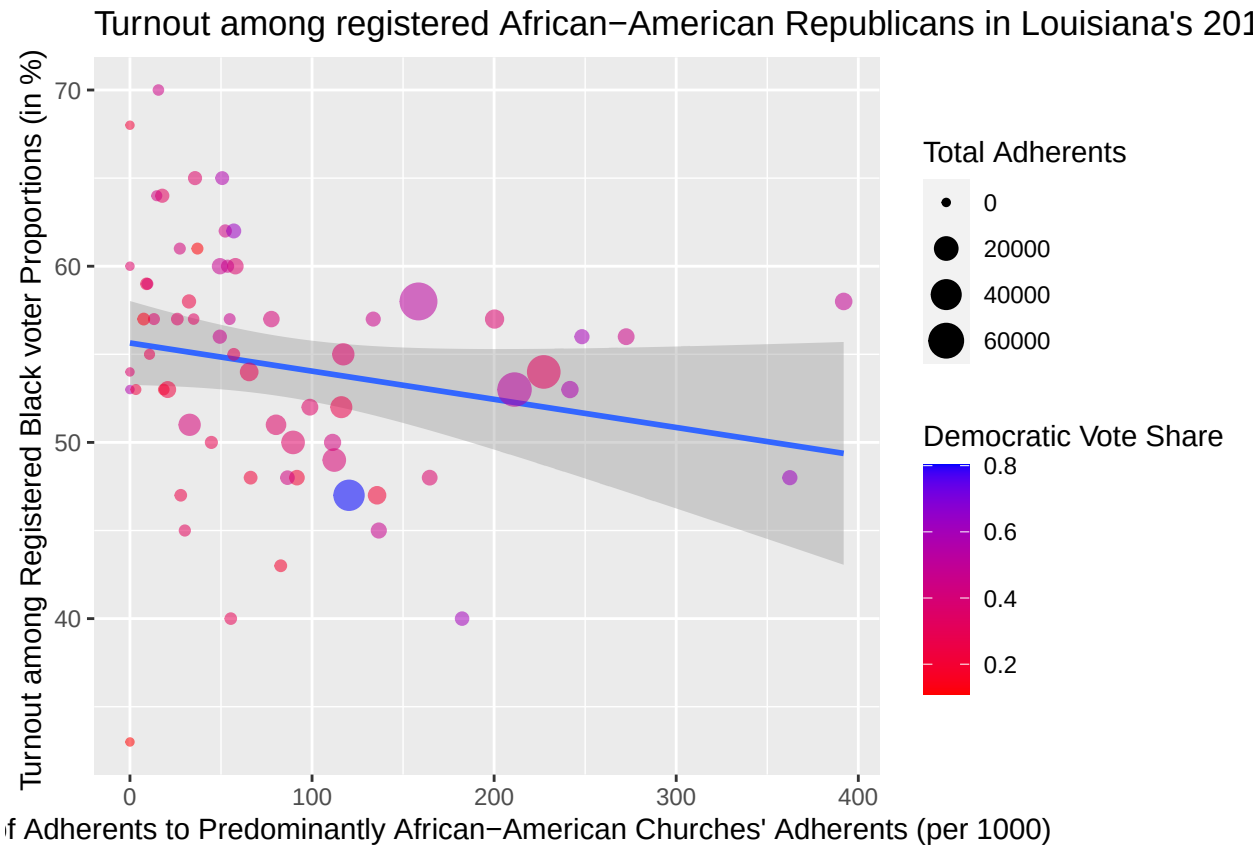
```
g_turnout_LA_2012_dem <- ggplot(data = Louisiana_general_AA, aes(x = AA_RATE, y = Black_dem_vote_prop))
g_turnout_LA_2012_dem +
  geom_smooth(method = "lm") +
  geom_point(aes(alpha = 0.05, color = dem_share, size = AA_ADH)) +
  guides(alpha = "none") +
  labs(x = "Rate of Adherents to Predominantly African-American Churches' Adherents (per 1000)", y = "Tu
  scale_color_gradient(low = "red", high = "blue")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



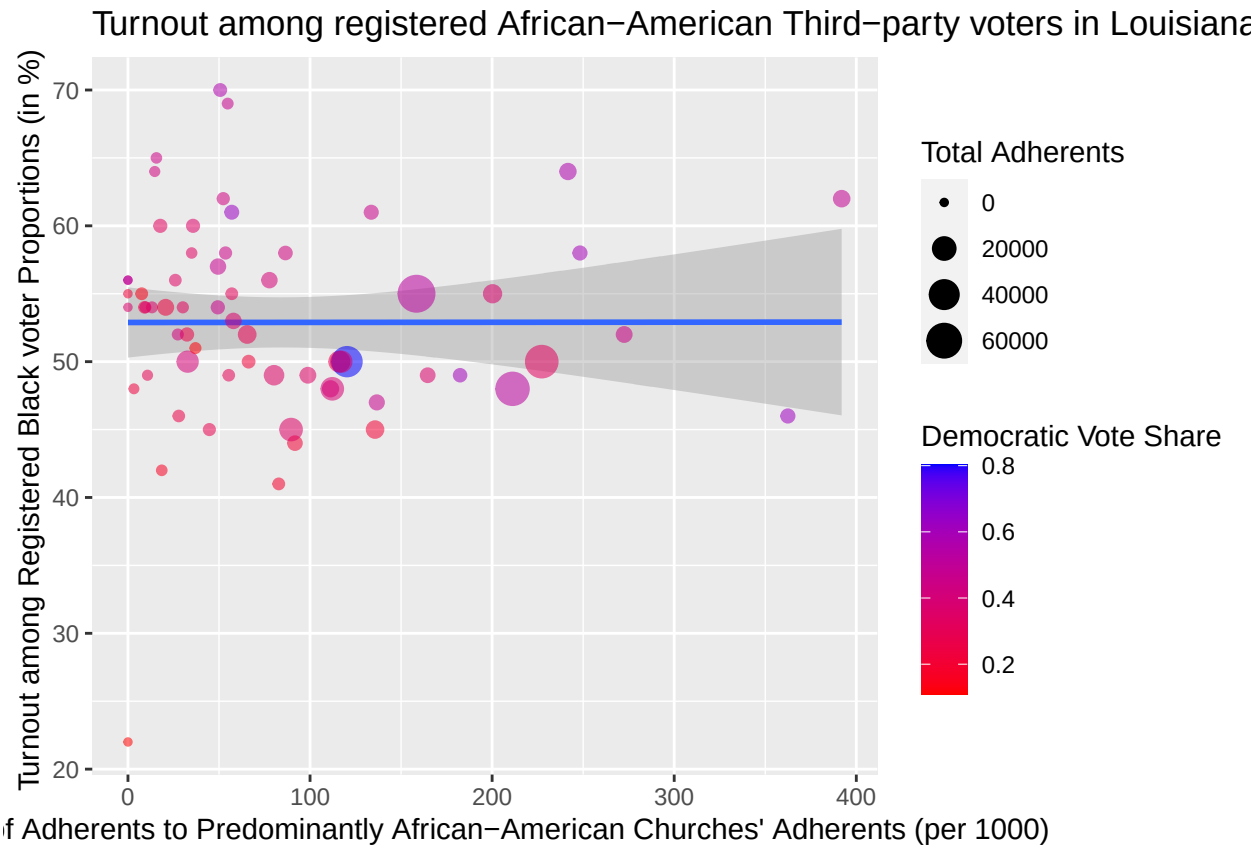

```
g_turnout_LA_2012_rep <- ggplot(data = Louisiana_general_AA, aes(x = AA_RATE, y = Black_rep_vote_prop))
g_turnout_LA_2012_rep +
  geom_smooth(method = "lm") +
  geom_point(aes(alpha = 0.05, color = dem_share, size = AA_ADH)) +
  guides(alpha = "none") +
  labs(x = "Rate of Adherents to Predominantly African-American Churches' Adherents (per 1000)", y = "Tu
  scale_color_gradient(low = "red", high = "blue")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
g_turnout_LA_2012_third <- ggplot(data = Louisiana_general_AA, aes(x = AA_RATE, y = Black_third_vote_pr
g_turnout_LA_2012_third +
  geom_smooth(method = "lm") +
  geom_point(aes(alpha = 0.05, color = dem_share, size = AA_ADH)) +
  guides(alpha = "none") +
  labs(x = "Rate of Adherents to Predominantly African-American Churches' Adherents (per 1000)", y = "Tu
  scale_color_gradient(low = "red", high = "blue")
```

```
## `geom_smooth()` using formula 'y ~ x'
```


Turnout among registered African−American Third−party voters in Louisiana
Turnout among Registered Black voter Proportions (in %)
70
60
50
40
30
20
0
100
200
300
400
f Adherents to Predominantly African−American Churches' Adherents (per 1000)
Total Adherents
0
20000
40000
60000
Democratic Vote Share
0.8
0.6
0.4
0.2