[redacted]

[redacted]

by

Scott Greenberg

Anumber: [redacted]

Date: [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

(i) **My Worst Color Choices Ever!** (25 Points):

You have to find one of your truly terrible, previously created R graphics with respect to colors. This should come from your current or past undergraduate, MS, or PhD research or from any of your HWs or projects for another class. Graphics you created for some outside employer or internship also are acceptable. Clearly indicate the source of this graphic. There must be at least five different "colors" in this graphic (other than black and white). If you claim that you never created any truly terrible graphic with respect to colors, find such a bad graphic on the web, e.g., Wikipedia, US Government web sites, etc. However, as a penalty, such an external graphic must have at least eight different "colors" (other than black and white).

You have to work on the following tasks:

(a) (5 Points) Show your original figure and describe what is bad with respect to your previously selected colors, based on your knowledge from the lecture materials.

<u>Answer:</u>

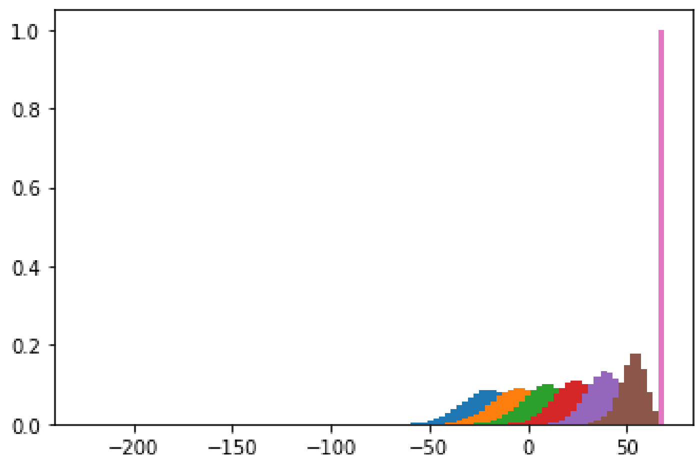
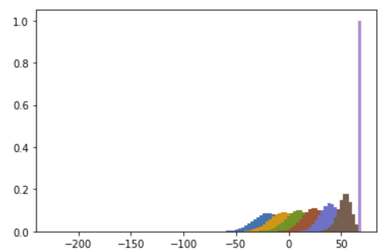


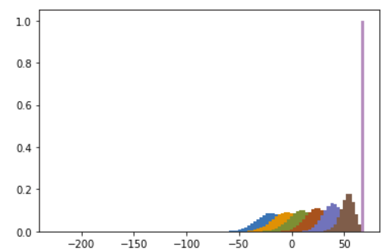
Figure 1: A bad graph I made for my internship. It was originally made in python, and was used to show how an increasing winrate changed the probability mass function. It is bad because it uses a qualitative scale for color when it would be best to use an ascending scale.

(b) (5 Points) Run your original figure through the eight color blindness simulation modes at `https://www.color-blindness.com/coblis-color-blindness-simulator/`. Save the results and include them as an answer to this question part. For which type of colorblindness (if any) did your original figure contain meaningful color information? Describe what is bad with respect to your previously selected colors, based on the results from the color blindness simulations.

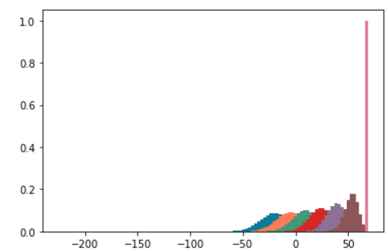
<u>Answer:</u>



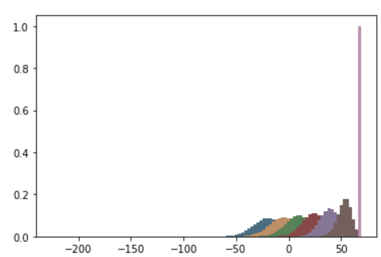
(a) Red-Weak/Protanomaly



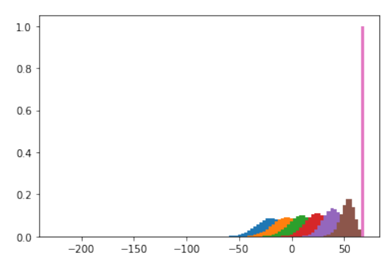
(b) Green-Weak/Deuteranomaly



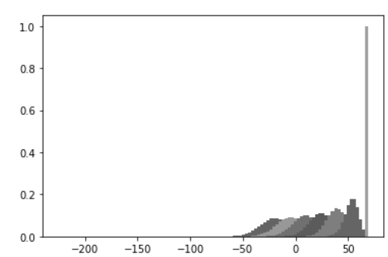
(c) Blue-Weak/Tritanomaly



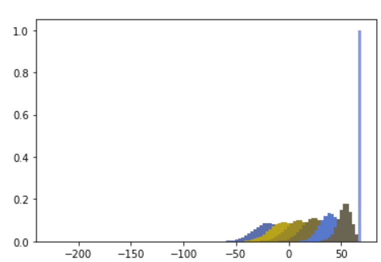
(a) Blue Cone Monochromacy



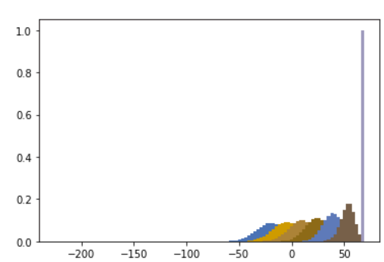
(b) Original Image



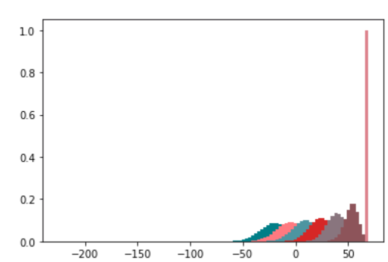
(c) Monochromacy/Achromatopsia



(a) Red-Blind/Protanopia



(b) Green-Blind/Deuteranopia



(c) Blue-Blind/Tritanopia

For which type of colorblindness (if any) did your original figure contain meaningful color information?

All of the types of colorblindness contain useful color information because with all of them, since one can tell the difference between adjacent pmf's and one can tell the difference between the start and ending pmf's, one can infer that the leftmost pmf is for .55 and the rightmost is for 1. However, if one was insistent not to infer anything about the graphs, the only colorblindnesses that are meaningful are: BLue-Weak/Tritanomaly, and Blue Cone Monochromacy (which is surprising since my understanding is that they are sort of opposites).
Describe what is bad with respect to your previously selected colors, based on the results from the color blindness simulations.

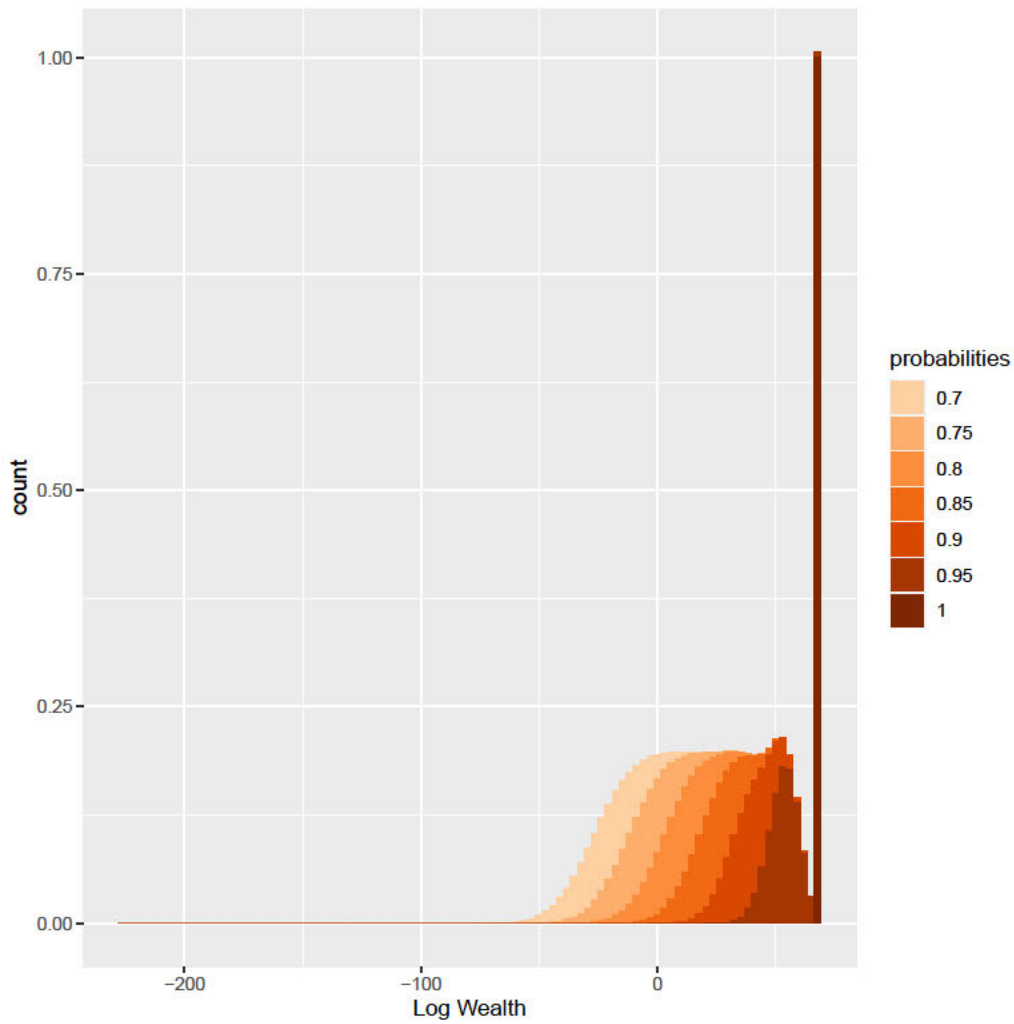
With the previously selected colors it is very hard/impossible to tell the difference between all the colors just based on the colors alone.

(c) (10 Points) Modify colors and any other design features (such as line styles, line thickness, plotting symbols, shading, etc.) from your original figure to make it more suitable for colorblind viewers. Check again with `https://www.color-blindness.com/coblis-color-blindness-simulator/`. Iterate this step until you have found a satisfactory solution. **Show your R code and include the final version of your figure and the final eight colorblindness simulations as an answer to this question part.**

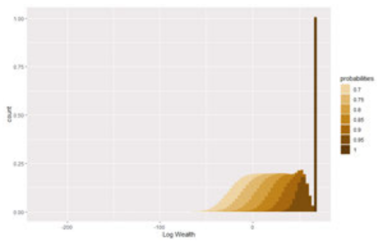
Be aware that there may be no perfect color choices for all of the simulated colorblindness scenarios. However, if your colors still remain hard to distinguish, your other design features should help to distinguish the different elements in your improved graph.

Answer:

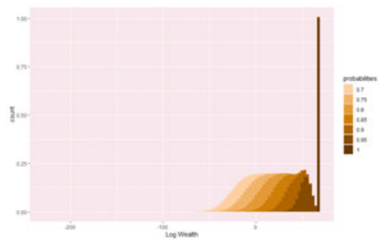
```
library(RColorBrewer)

## Warning:  package 'RColorBrewer' was built under R version 4.0.3

library(ggplot2)

## Warning:  package 'ggplot2' was built under R version 4.0.3

wealth.plus.prob <- cbind(read.csv('WTExpl.csv')[, -1],
                          read.csv('pWTExpl.csv')[, -1])
colnames(wealth.plus.prob) <- c('Wealth',
                                read.csv('pExpl.csv')$X0)
oranges <- brewer.pal(9, 'Oranges')[-1:-2]
hist <- cbind(rep(wealth.plus.prob$Wealth, times = 7),
              unlist(lapply(read.csv('pExpl.csv')$X0, rep,
                            times = length(wealth.plus.prob$Wealth))),
              as.vector(unlist(wealth.plus.prob[, 2:8])))
ggplot() + geom_histogram(aes(x = log(hist[, 1]), weight = hist[, 3],
                              group = as.character(hist[, 2]),
                              fill = as.character(hist[, 2])),
                          bins = length(wealth.plus.prob$Wealth)) +
scale_fill_manual(name = 'probabilities', values = oranges) +
xlab('Log Wealth')
```


count
1.00
0.75
0.50
0.25
0.00
−200
−100
0
Log Wealth
probabilities
0.7
0.75
0.8
0.85
0.9
0.95
1

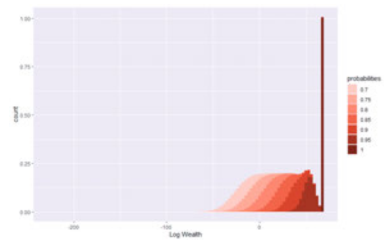
Comments:



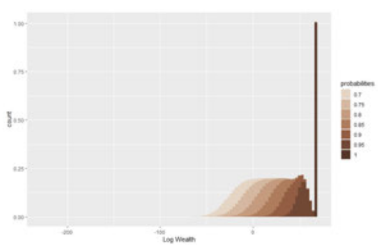
(a) Red-Weak/Protanomaly



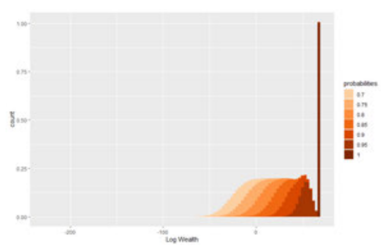
(b) Green-Weak/Deuteranomaly



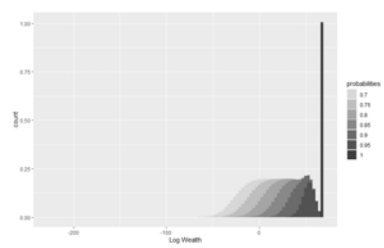
(c) Blue-Weak/Tritanomaly



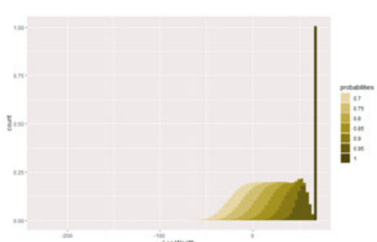
(a) Blue Cone Monochromacy



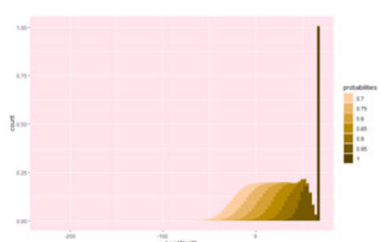
(b) Original Image



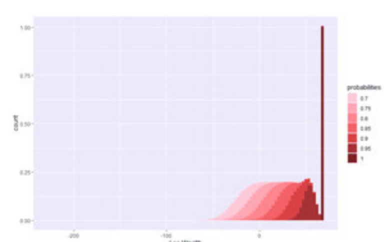
(c) Monochromacy/Achromatopsia



(a) Red-Blind/Protanopia



(b) Green-Blind/Deuteranopia



(c) Blue-Blind/Tritanopia

(d) (5 Points) Clearly describe how you have modified your figure to make it more appropriate for colorblind viewers. Mention the names of the color palette you used, how many colors you used, whether you omitted some colors from this palette (e.g., just used 4 colors from a 5–color palette), etc. Likely, it was not enough to just choose a colorblind color palette from *RColorBrewer*. Rather, additional changes may have been necessary. Do not forget to mention these modifications as well.

Answer:

I modified my figure by choosing a colorblind safe palette. The palette I used was the 9-color oranges, except I omitted the two lightest shades (#fff5eb, and #fee6ce) since they blended with the ggplot background, especially in Monochromacy/Achromatopsia, and Blue Cone Monochromacy.

[redacted]

(a) (4 Points) Find a data set that is of interest to you with **two related variables** for the 50 US states. Washington, DC, is optional. This can be disease data, economic data, unemployment data, climate data, or any other data set that is of interest to you. **Related** means that both variables are on the same scale and have the same units, e.g., Covid–19 rates in January 2021 and Covid–19 rates in November 2020 in each of the 50 states, or unemployment rates on February 28, 2020, and April 30, 2020, in each of the 50 states. The data set must have been published in the past 4 years (2017–2021). Present your proposed data set to me via e–mail by **Monday 3/22/2021, 11:59pm**. Be specific about the source of your data set and include the URL or other bibliographic information.

Provide some basic summary and the source/URL of your data set here as well. Also show the content of this data set so I can check your maps for correctness. After all, your data set should only have 50 (or 51) rows and three columns (states and two variables).

Answer:

```
library(openxlsx)

## Warning:  package 'openxlsx' was built under R version 4.0.4

n.t <- read.xlsx(paste0('https://www2.census.gov/programs-surveys/cp',
                        's/tables/time-series/historical-poverty-peo',
                        'ple/hstpov19.xlsx'))
n.t.r <- data.frame(matrix(nrow = 51, ncol = 3))
n.t.r[, 1] <- n.t[6:56, 1]
n.t.r[, 2] <- n.t[6:56, 2]
n.t.r[, 3] <- n.t[6:56, 4]

colnames(n.t.r) <- c(n.t[3, 1], paste(n.t[3, 2], n.t[4, 2]),
                     paste(n.t[3, 4], n.t[4, 4]))
n.t.r

##                    State      2019 Percent      2018 Percent
## 1                Alabama              12.9                16
## 2                 Alaska              10.2              13.1
```

```
## 3               Arizona              9.9              12.8
## 4              Arkansas             14.1              15.9
## 5            California             10.1              11.9
## 6              Colorado 9.300000000000001               9.1
## 7           Connecticut 8.300000000000001              10.2
## 8              Delaware              6.5               7.4
## 9  District of Columbia             12.5              14.7
## 10              Florida             11.5              13.7
## 11              Georgia             12.1              14.8
## 12               Hawaii              8.4 9.199999999999999
## 13                Idaho              7.1              11.5
## 14             Illinois 9.300000000000001              10.3
## 15              Indiana             10.1              11.6
## 16                 Iowa              9.5               8.9
## 17               Kansas              9.5               7.5
## 18             Kentucky             13.6              15.7
## 19            Louisiana             17.9                19
## 20                Maine             10.4              11.6
## 21             Maryland                7                 8
## 22        Massachusetts              7.5 8.699999999999999
## 23             Michigan             10.2              10.5
## 24            Minnesota              5.7               7.9
## 25          Mississippi             19.2              19.6
## 26             Missouri              9.4              12.4
## 27              Montana 9.699999999999999              10.3
## 28             Nebraska 8.699999999999999              10.5
## 29               Nevada             10.4                13
## 30        New Hampshire              3.7               6.1
## 31           New Jersey              6.3 8.199999999999999
## 32           New Mexico             15.3              16.6
## 33             New York             12.5              11.1
## 34       North Carolina             12.7              13.1
## 35         North Dakota              8.1 9.699999999999999
## 36                 Ohio             12.4              11.9
## 37             Oklahoma             10.8              13.4
## 38               Oregon              8.1 9.699999999999999
## 39         Pennsylvania 8.699999999999999              11.8
## 40         Rhode Island 9.199999999999999               8.9
## 41       South Carolina             15.1              12.8
## 42         South Dakota             10.6              10.6
## 43            Tennessee             13.1                12
## 44                Texas             11.1              13.7
## 45                 Utah              7.3               6.9
## 46              Vermont              8.6 9.699999999999999
## 47             Virginia 8.800000000000001 9.800000000000001
## 48           Washington                7               8.6
## 49        West Virginia             13.9              15.9
## 50            Wisconsin              8.4               8.6
## 51              Wyoming 9.199999999999999               9.5
```

Comments:

2018&2019 state poverty rates: https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-poverty-people/hstpov19.xlsx

(b) (8 Points) To better understand your data set, first create two choropleth maps, one for each variable. Use a map that either includes Alaska and Hawaii, or be sure to add them in as was demonstrated in the lecture materials. As your two variables are related, use a common sequential or divergent color scheme with meaningful identical intervals, say [1%, 2%), [2%, 3%), etc. Make sure to use the same intervals for both maps. Double check that the results you get match the results you expect from the given data.

Answer:

```
library(maps)

## Warning:  package 'maps' was built under R version 4.0.4

library(mapdata)

## Warning:  package 'mapdata' was built under R version 4.0.4

library(classInt)

## Warning:  package 'classInt' was built under R version 4.0.4

n.t.r[, 2:3] <- as.numeric(unlist(n.t.r[, 2:3]))

n.t.r.b <- c(n.t.r[, 2], n.t.r[, 3])

breaks <- classIntervals(n.t.r.b, n = 5, style = "quantile",
                         intervalClosure = "right")

r.y.b <- rev(brewer.pal(5, 'RdYlBu'))

m.class <- cut(n.t.r[, 2], breaks$brks, include.lowest = TRUE)
m.col <- r.y.b[m.class]
map.m.col <- m.col[match.map('state', n.t.r[, 1])]

pdf('hw03_poverty2019.pdf', width = 7, height = 3)
layout(matrix(c(2, 2, 1, 1, 1, 1, 1,
                2, 2, 1, 1, 1, 1, 1,
                2, 2, 1, 1, 1, 1, 4,
                3, 3, 1, 1, 1, 1, 4,
                3, 3, 1, 1, 1, 1, 1,
                3, 3, 1, 1, 1, 1, 1), 6, 7,
              byrow = TRUE))

par(mar = c(0, 0, 3, 0))
map('state', fill = TRUE, col = map.m.col,  interior = TRUE)
legend('bottomright',
       title = 'Poverty rate (Percentages)',
       legend = levels(m.class),
       cex = 0.6,
       fill = r.y.b,
       bty = 'n')
title('Poverty rate by states in 2019')
```

```r
par(mar = c(0, 0, 0, 0))
map('world2Hires', 'USA:Alaska',
    fill = TRUE,
    col = r.y.b[m.class[n.t.r[, 1] == 'Alaska']])

par(mar = c(0, 0, 0, 0))
map('world2Hires', 'Hawaii',
    xlim = c(199, 206),
    ylim = c(18.5, 23),
    fill = TRUE,
    col = r.y.b[m.class[n.t.r[, 1] == 'Hawaii']])

par(mar = c(0, 0, 0, 0))
map('county', 'district of columbia,washington',
    fill = TRUE,
    col = r.y.b[m.class[n.t.r[, 1] == 'District of Columbia']])

dev.off()

## pdf
##   2

breaks2 <- classIntervals(n.t.r.b, n = 5, style = "quantile",
                          intervalClosure = "right")


m.class2 <- cut(n.t.r[, 3], breaks2$brks, include.lowest = TRUE)
m.col2 <- r.y.b[m.class2]
map.m.col2 <- m.col2[match.map('state', n.t.r[, 1])]

pdf('hw03_poverty2018.pdf', width = 7, height = 3)
layout(matrix(c(2, 2, 1, 1, 1, 1, 1,
                2, 2, 1, 1, 1, 1, 1,
                2, 2, 1, 1, 1, 1, 4,
                3, 3, 1, 1, 1, 1, 4,
                3, 3, 1, 1, 1, 1, 1,
                3, 3, 1, 1, 1, 1, 1), 6, 7,
              byrow = TRUE))

par(mar = c(0, 0, 3, 0))
map('state', fill = TRUE, col = map.m.col2,  interior = TRUE)
legend('bottomright',
       title = 'Poverty rate (Percentages)',
       legend = levels(m.class2),
       cex = 0.6,
       fill = r.y.b,
       bty = 'n')
title('Poverty rate by states in 2018')

par(mar = c(0, 0, 0, 0))
map('world2Hires', 'USA:Alaska',
    fill = TRUE,
    col = r.y.b[m.class2[n.t.r[, 1] == 'Alaska']])
```

```r
par(mar = c(0, 0, 0, 0))
map('world2Hires', 'Hawaii',
    xlim = c(199, 206),
    ylim = c(18.5, 23),
    fill = TRUE,
    col = r.y.b[m.class2[n.t.r[, 1] == 'Hawaii']])


par(mar = c(0, 0, 0, 0))
map('county', 'district of columbia,washington',
    fill = TRUE,
    col = r.y.b[m.class2[n.t.r[, 1] == 'District of Columbia']])

dev.off()

## pdf
##   2
```

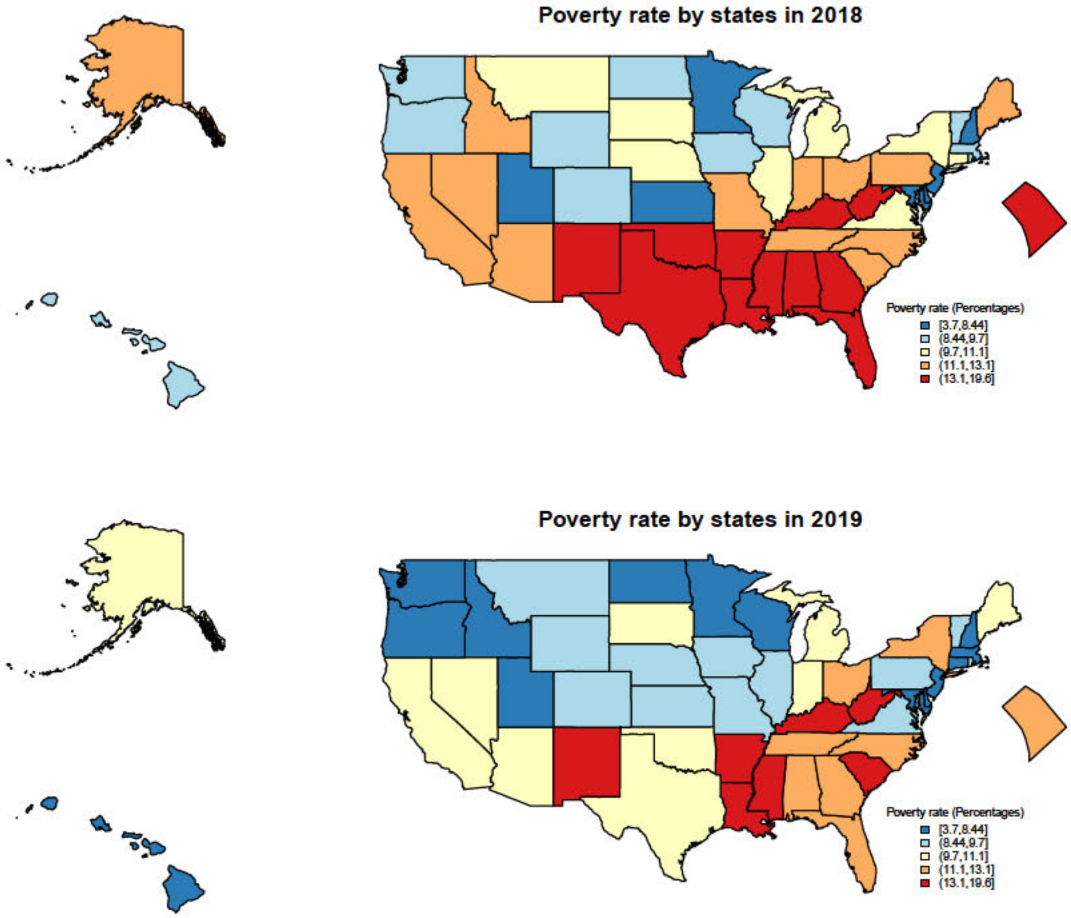


Figure 8: Four choropleth maps showing the poverty rates by state. The map at the top is of 2019. The map at the bottom is of 2018.

(c) (4 Points) Write about 5 or 6 sentences about your observations when comparing these two maps. Are there any strong regional patterns in your two maps? Where? Are there any major visible differences for some (all) regions in your two maps? Be specific.

Answer:

The only pattern I noticed is that place with very high poverty rates in 2018 usually have very high poverty rates in 2019. I have not noticed any regional patterns or differences, except the southeast usually has higher poverty rates.

(d) (8 Points) In this part, you have to create comparative micromaps that depict the changes between the two original maps. To create these maps, you may want to create a new column in the data structure that shows the change from the "base" map to the "follow-up" map. For two temporal variables, the "base" map should show the earlier point in time. For all other variables, you have to decide what is the "base" map and what is the "follow–up" map and then calculate differences accordingly.

There should be four maps overall: One map will show the states that experienced an increase from the "base" map to the "follow–up" map. One map will show the states that experienced a decrease from the "base" map to the "follow-up" map. You should also reuse the two choropleth maps from part (b). n.t. The layout of these four choropleth maps should be as follows: The map showing increases should be centered in the top row, the two maps from part (b) should be shown side by side ("base" map on the left and "follow–up" map on the right) in the center row, and the map showing decreases should be centered in the bottom row. Look at Figure 56 from the pdf lecture notes, but keep in mind that your figures only needs to contain four maps and not 25 maps as in that example.

Include legends and titles for each map separately. These maps should appear as one figure. I would suggest writing the four maps to a separate pdf file each, showing your R code and evaluating your R code, but suppressing the figure inside the R chunk and rather including the maps in the LaTeX document immediately following the code chunk (google for `latex multiple figures together`; there exist numerous options how to combine multiple figures within LaTeX). Using `layout` in R should work as well, but may be more complicated overall.

<u>Answer:</u>

```
library(assertthat)

## Warning:  package 'assertthat' was built under R version 4.0.3

is.equal <- mapply(are_equal, m.class, m.class2)
l.t.n <- n.t.r[is.equal == FALSE, 2] < n.t.r[is.equal == FALSE, 3]
l.t.n.s <- n.t.r[is.equal == FALSE, ][l.t.n == TRUE, ]
g.t.n.s <- n.t.r[is.equal == FALSE, ][l.t.n == FALSE, ]

breaks3 <- classIntervals(n.t.r.b, n = 5, style = "quantile",
                          intervalClosure = "right")
```

```r
m.class3 <- cut(l.t.n.s[, 2], breaks3$brks, include.lowest = TRUE)
m.col3 <- r.y.b[m.class3]
map.m.col3 <- m.col3[match.map('state', l.t.n.s[, 1])]

pdf('hw03_lowerpoverty2019.pdf', width = 7, height = 3)
layout(matrix(c(2, 2, 1, 1, 1, 1, 1,
                2, 2, 1, 1, 1, 1, 1,
                2, 2, 1, 1, 1, 1, 4,
                3, 3, 1, 1, 1, 1, 4,
                3, 3, 1, 1, 1, 1, 1,
                3, 3, 1, 1, 1, 1, 1), 6, 7,
              byrow = TRUE))

par(mar = c(0, 0, 3, 0))
map('state', fill = TRUE, col = map.m.col3,  interior = TRUE)
legend('bottomright',
       title = 'Poverty rates (Percentages)',
       legend = levels(m.class3),
       cex = 0.6,
       fill = r.y.b,
       bty = 'n')
title(paste('States experiencing a decrease in Poverty Rates from',
            '2018 to 2019'))
par(mar = c(0, 0, 0, 0))
map('world2Hires', 'USA:Alaska',
    fill = TRUE,
    col = r.y.b[m.class3[l.t.n.s == 'Alaska']])

par(mar = c(0, 0, 0, 0))
map('world2Hires', 'Hawaii',
    xlim = c(199, 206),
    ylim = c(18.5, 23),
    fill = TRUE,
    col = r.y.b[m.class3[l.t.n.s == 'Hawaii']])

par(mar = c(0, 0, 0, 0))
map('county', 'district of columbia,washington',
    fill = TRUE,
    col = r.y.b[m.class3[l.t.n.s == 'District of Columbia']])

dev.off()

## pdf
##   2

breaks4 <- classIntervals(n.t.r.b, n = 5, style = "quantile",
                          intervalClosure = "right")


m.class4 <- cut(g.t.n.s[, 2], breaks4$brks, include.lowest = TRUE)
m.col4 <- r.y.b[m.class4]
map.m.col4 <- m.col4[match.map('state', g.t.n.s[, 1])]

pdf('hw03_higherpoverty2019.pdf', width = 7, height = 3)
```

```r
layout(matrix(c(2, 2, 1, 1, 1, 1, 1,
                2, 2, 1, 1, 1, 1, 1,
                2, 2, 1, 1, 1, 1, 4,
                3, 3, 1, 1, 1, 1, 4,
                3, 3, 1, 1, 1, 1, 1,
                3, 3, 1, 1, 1, 1, 1), 6, 7,
              byrow = TRUE))

par(mar = c(0, 0, 3, 0))
map('state', fill = TRUE, col = map.m.col4,  interior = TRUE)
legend('bottomright',
       title = 'Poverty rate (Percentages)',
       legend = levels(m.class4),
       cex = 0.6,
       fill = r.y.b,
       bty = 'n')
title(paste('States experiencing an increase in Poverty Rates from',
            '2018 to 2019'))
par(mar = c(0, 0, 0, 0))
map('world2Hires', 'USA:Alaska',
    fill = TRUE,
    col = r.y.b[m.class4[g.t.n.s == 'Alaska']])

par(mar = c(0, 0, 0, 0))
map('world2Hires', 'Hawaii',
    xlim = c(199, 206),
    ylim = c(18.5, 23),
    fill = TRUE,
    col = r.y.b[m.class4[g.t.n.s == 'Hawaii']])


par(mar = c(0, 0, 0, 0))
map('county', 'district of columbia,washington',
    fill = TRUE,
    col = r.y.b[m.class4[g.t.n.s == 'District of Columbia']])

dev.off()

## pdf
##   2
```

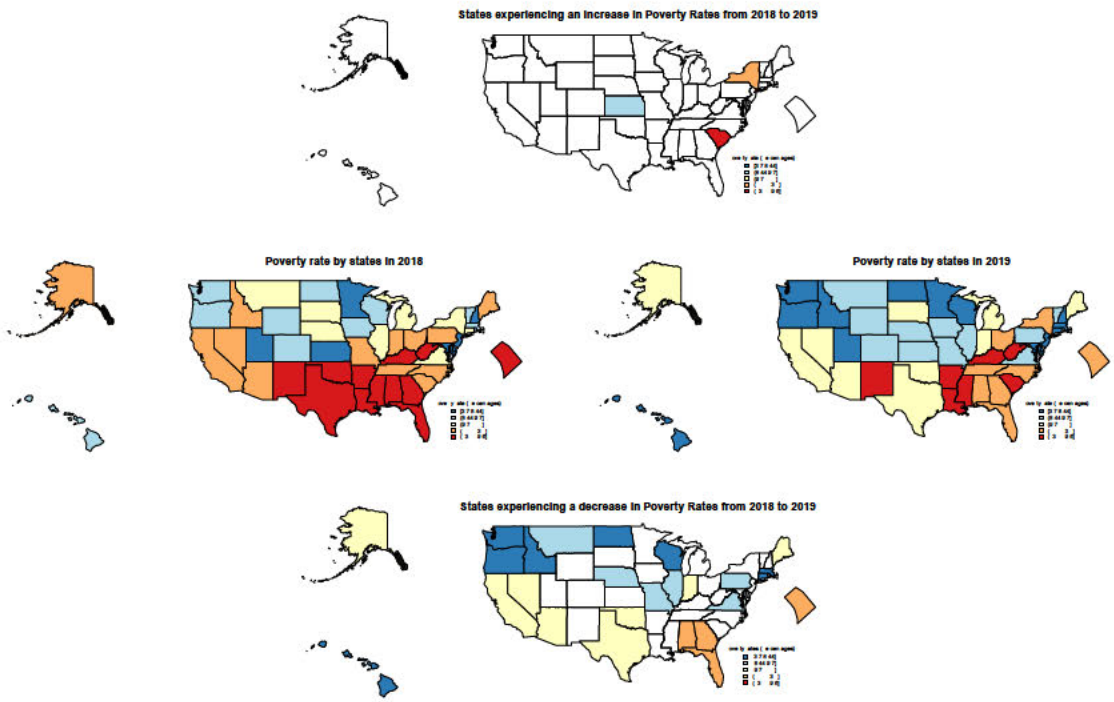


Figure 9: Two choropleth maps showing the poverty rates by state. The map at the right is of 2019. The map at the left is of 2018. The map at the top shows states that increased to a higher category. The map at the bottom shows states that decreased to a lower category.

(e) (4 Points) Again, write about 5 or 6 sentences about your observations when comparing these four maps. Are there any strong regional change patterns in your two comparative maps? Where? Are there any major visible differences for some (all) regions in your two comparative maps? Be specific. What can you see in the additional comparative maps that was not immediately visible in your two choropleth maps from part (b)?

Answer:

There is a greater amount of decrease in poverty rate, than an increase in poverty rate. I notice no additional regional patterns in my two comparative maps except that states in a lower category of poverty rates for 2019 are in clusters. With these comparative maps the main thing I see now that was not immediately visible before is a lack of change in many states' poverty rate.

(f) (4 Points) Create the required data structures (such as the map table) and then draw a basic linked micromap plot for your data set, based on code similar to the five to eight line–basic examples we used in class and in the micromap workshop materials. Just show dotplots of the data in two statistical panels. Include a figure of your basic linked micromap plot, ideally written to an external file and then incorporated back into R.

Answer:

```
library(micromap)

## Warning:  package 'micromap' was built under R version 4.0.4
## Loading required package:  maptools
## Warning:  package 'maptools' was built under R version 4.0.4
## Loading required package:  sp
## Warning:  package 'sp' was built under R version 4.0.4
## Checking rgeos availability:  FALSE
##  Note:  when rgeos is not available, polygon geometry computations in maptools depend on gpclib,
##  which has a restricted licence.  It is disabled by default;
##  to enable gpclib, type gpclibPermit()
## Loading required package:  rgdal
## Warning:  package 'rgdal' was built under R version 4.0.4
## rgdal:  version:  1.5-23, (SVN revision 1121)
## Geospatial Data Abstraction Library extensions to R successfully loaded
## Loaded GDAL runtime:  GDAL 3.2.1, released 2020/12/29
## Path to GDAL shared files:  C:/Users/Scott/Documents/R/win-library/4.0/rgdal/gdal
## GDAL binary built with GEOS: TRUE
## Loaded PROJ runtime:  Rel.  7.2.1, January 1st, 2021, [PJ_VERSION: 721]
## Path to PROJ shared files:  C:/Users/Scott/Documents/R/win-library/4.0/rgdal/proj
## PROJ CDN enabled:  FALSE
## Linking to sp version:1.4-5
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,
## use options("rgdal_show_exportToProj4_warnings"="none") before loading rgdal.
## Overwritten PROJ_LIB was C:/Users/Scott/Documents/R/win-library/4.0/rgdal/proj

library(labeling)

## Warning:  package 'labeling' was built under R version 4.0.3

data("USstates")

statePolys <- create_map_table(USstates, 'ST_NAME')
n.t.r.lm <- n.t.r
n.t.r.lm[n.t.r.lm[, 1] == 'District of Columbia',
         1] <- as.vector(statePolys$ID[64])[1]
pdf('hw03_basiclm.pdf', width = 8.25, height = 10)

mmplot(stat.data = n.t.r.lm,
       map.data = statePolys,
       map.link = c('State', 'ID'),
       panel.types = c('dot_legend', 'labels', 'dot', 'dot', 'map'),
```

```r
       panel.data = list(NA, 'State', '2018 Percent', '2019 Percent',
                         NA),
       ord.by = '2019 Percent',
       grouping = 5, median.row = TRUE)
dev.off()

## pdf
##   2
```

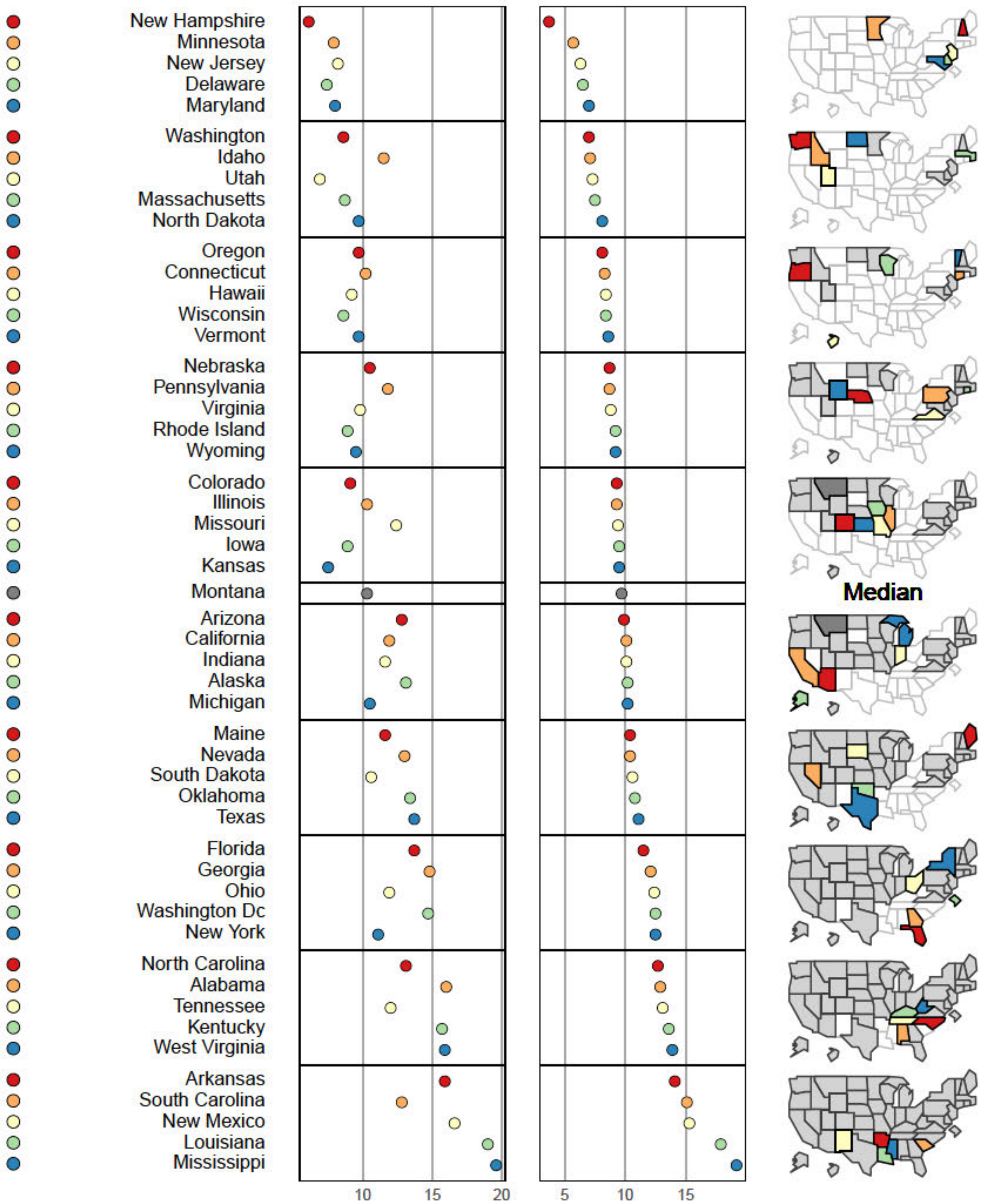


Figure 10: Linked Micromap for the states.

(g) (8 Points) Optimize your basic linked micromap plot and make it ready for publication. Follow improvements from class and from the micromap workshop materials, such as changing colors, adjusting axis labels, adding reference lines (for national means or medians), etc. Include a figure of your final linked micromap plot, ideally written to an external file and then incorporated back into R.

<u>Answer:</u>

```
pdf('hw03_advancedlm.pdf', width = 8.25, height = 10)
tl2018 <- as.list(extended(min(n.t.r.lm[, 3]), max(n.t.r.lm[, 3]), 4))
tl2019 <- as.list(extended(min(n.t.r.lm[, 2]), max(n.t.r.lm[, 2]), 4))
mmplot(stat.data = n.t.r.lm,
       map.data = statePolys,
       map.link = c('State', 'ID'),
       panel.types = c('labels', 'dot_legend', 'map', 'dot', 'dot'),
       panel.data = list('State', NA, NA, '2018 Percent',
                         '2019 Percent'),
       ord.by = '2019 Percent',
       vertical.align='center',
       colors=oranges[c(-1, -3)],
       grouping = 5, median.row = TRUE,
       panel.att = list(list(1, header = 'States', panel.width =  0.8,
                             text.size = 0.9),

                        list(2, point.type = 20, point.border = TRUE,
                             point.size = 2),

                        list(3,
                             header = paste('Light gray is the',
                                            '\npreviously displayed'),
                             map.all = TRUE,
                             fill.regions ='aggregate',
                             active.border.color = 'black',
                             active.border.size = 0.75,
                             inactive.border.color = gray(.7),
                             inactive.border.size = 0.5,
                             panel.width = .8),

                        list(4,
                             header = '2018 Poverty Rate (Percent)',
                             point.size = 1,
                             xaxis.ticks = tl2018,
                             xaxis.labels = tl2018,
                             xaxis.title = 'Percent'),

                        list(5,
                             header = '2019 Poverty Rate (Percent)',
                             point.size = 1,
                             xaxis.ticks = tl2019,
                             xaxis.labels = tl2019,
```

```
                              xaxis.title = 'Percent')))
dev.off()

## pdf
##   2
```

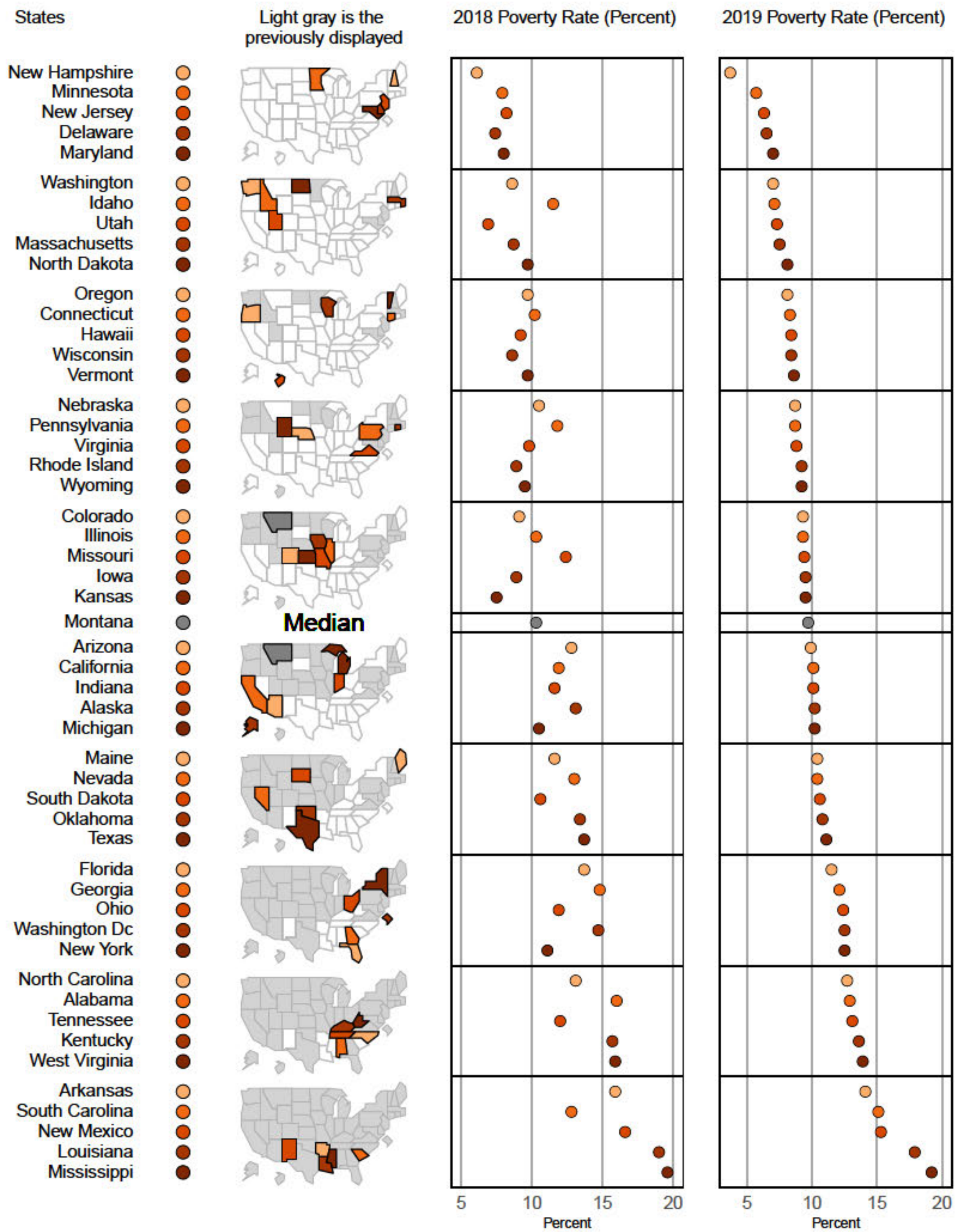


Figure 11: Improved Linked Micromap for the States.

(h) (4 Points) Once more, write about 5 or 6 sentences about your observations when comparing these two variables via a linked micromap plot. Comment on any regional pattern of your sorting variable, make a crude assessment of the correlation between the two variables in your linked micromap plot (and numerically verify this assessment). Report on any noticeable outliers — both spatially in the maps and in the two statistical columns (recall how we discussed ouliers for the soybeans linked micromap plot in class).

<u>Answer:</u>

States with similar poverty rates are sort of clustered together in small groups. In addition, in 2019 some states have very low poverty rates like New Hampshire, that seem signifanctly lower than all other states; while some, like Mississippi and Louisiana, have a significantly higher percentage than any other state. Also, for 2019, the percentages for states that are moderately high to very high in poverty rates are spread farther apart than the states with moderately low to very low poverty rates. However, there is no geographic clustering which is possibly due to something like population or the geographic distance between metropolitan areas in the United States.

(i) (6 Points) Write about 1/2 to 1 page to compare the results you obtained in parts (b), (d), and (g). What can be best seen in the choropleth maps, what do we gain from the comparative micromaps (if anything at all), and what do we learn from the linked micromap plot (if anything at all)?

Your answer likely will differ from answers given by other students, depending on whether big differences between your two variables occur in large states such as Texas or California or in small states such as Delaware and Rhode Island.

<u>Answer:</u>

In part (b) we see a southern cluster of states with high poverty rates back in 2018 (cluster of red). From part (d) we see that a large amount of states decreased there poverty rate category from 2018 to 2019. Finally in part (g) we see the how the correlation between 2018 rates and 2019 rates, as well as the geographic location of groups (not necessarily by geography) of states with similar poverty ranking.

(j) (+5 EC Points) **Extra Credit:**

Create a new linked micromap plot. Rather than including two panels showing your two variables, use one panel that shows the differences between the data used for the "base" map and the data used for the "follow–up" map in part (d). This is called an arrow plot and a detailed walk through on implementing this plot can be found on pages 23–33 of the Micromap User Guide (`R_Package_MicromapUserGuide_03_2014.pdf`, available in Canvas).

Do not include a median row. The arrow should point in the direction of change from the "base" map data to the "follow–up" map data (meaning the arrow base shows the "base" map data and the arrow tip the "follow–up" map data for that state). Order the rows of this linked micromap plot by the "follow–up" map data where the top of the linked micromap plot shows the high values of the "follow–up" map data.

Answer:

```
# Place your answer here
```

General Instructions

(i) Create a single pdf document, using R Markdown, Sweave, or knitr. When you take this course at the 6000–level, you have to use LaTeX in combination with Sweave or knitr. You only have to submit this one document to Canvas.

(ii) Include a title page that contains your name, your A–number, the number of the assignment, the submission date, and any other relevant information.

(iii) Start your answers to each main question on a new page (continuing with the next part of a question on the same page is fine). Clearly label each question and question part. Your answer to question (i) should start on page 2!

(iv) Show your R code and resulting graph(s) [if any] for each question part!

(v) Before you submit your homework, check that you follow all recommendations from Google's R Style Guide (see `http://web.stanford.edu/class/cs109l/unrestricted/resources/google-style.html`). Moreover, make sure that your R code is consistent, i.e., that you use the same type of assignments and the same type of quotes throughout your entire homework.

(vi) Give credit to external sources, such as stackoverflow or help pages. Be specific and include the full URL where you found the help (or from which help page you got the information). Consider R code from such sources as "legacy code or third–party code" that does not have to be adjusted to Google's R Style (even though it would be nice, in particular if you only used a brief code segment).

(vii) **Not following the general instructions outlined above will result in point deductions!**

(viii) For general questions related to this homework, please use the corresponding discussion board in Canvas! I will try to reply as quickly as possible. Moreover, if one of you knows an answer, please post it. It is fine to refer to web pages and R commands, but do not provide the exact R command with all required arguments or which of the suggestions from a stackoverflow web page eventually worked for you! This will be the task for each individual student!

(ix) Submit your single pdf file via Canvas by the submission deadline. Late submissions will result in point deductions as outlined on the syllabus.

1 References

https://www.dummies.com/programming/r/how-to-repeat-vectors-in-r/

https://stackoverflow.com/a/39322348

https://www.rdocumentation.org/packages/openxlsx/versions/4.2.3/topics/read.xlsx

https://statisticsglobe.com/change-legend-title-ggplot-r

https://statisticsglobe.com/coerce-list-object-to-type-double-in-r

https://stackoverflow.com/questions/18933441/how-to-reverse-order-a-vector