

S&DS 363 Homework 2

Kelly Farley + Evan Collins + Ken Stier

25 February 2021

The Dataset

COVID-19 infection and death statistics from U.S. counties (sourced from NYT), combined with economic, education, and population data (sourced from various government agencies) and also survey responses about mask-wearing frequencies (sourced from NYT). 3141 complete observations on 19 metric variables, with the first variable being a categorical county identifier (FIPS). To avoid any outliers due to population size differences between counties, all variables are scaled as a percentage of population. Variable descriptions can be found [here](#).

```
url_data = ("https://evancollins.com/covid_and_demographics.csv")
raw <- read_csv(url(url_data))

## Warning: Missing column names filled in: 'X1' [1]

##
## -- Column specification -----
## cols(
##   .default = col_double(),
##   X1 = col_character(),
##   County_Name = col_character(),
##   State_Name = col_character(),
##   FIPS = col_character()
## )
## i Use `spec()` for the full column specifications.

raw <- as.data.frame(raw)
db <- subset(raw, select=c(4:25)) # exclude ID columns except FIPS
db <- subset(db, select=-c(17,18)) # exclude categorical rural and economic codes
dim(db)

## [1] 3141    20
```

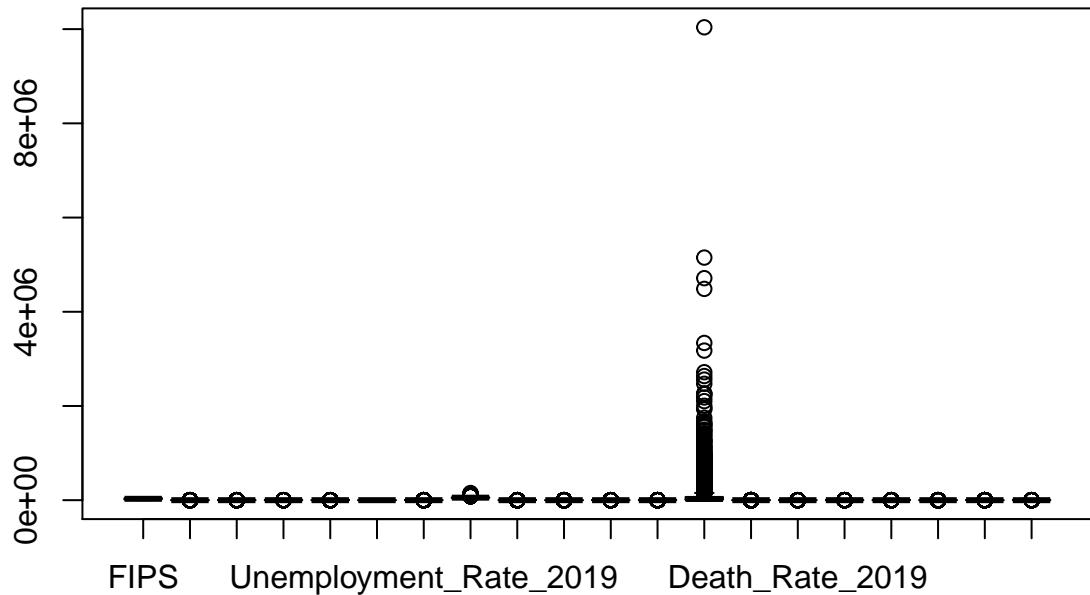
1: Multivariate Normal Distribution

```

# names of variables
names <- names(db)

# boxplot of all variables
db$FIPS <- as.numeric(db$FIPS) # make county codes numeric
boxplot(db)

```



```

# from examination of boxplots, col 8, 9, 10, 13 require transformation
trans <- c(8, 9, 10, 13)
for(i in 1:length(trans)){
  db[[paste("log", names[trans[i]])]] <- log(db[[names[trans[i]]]] + .01)
}

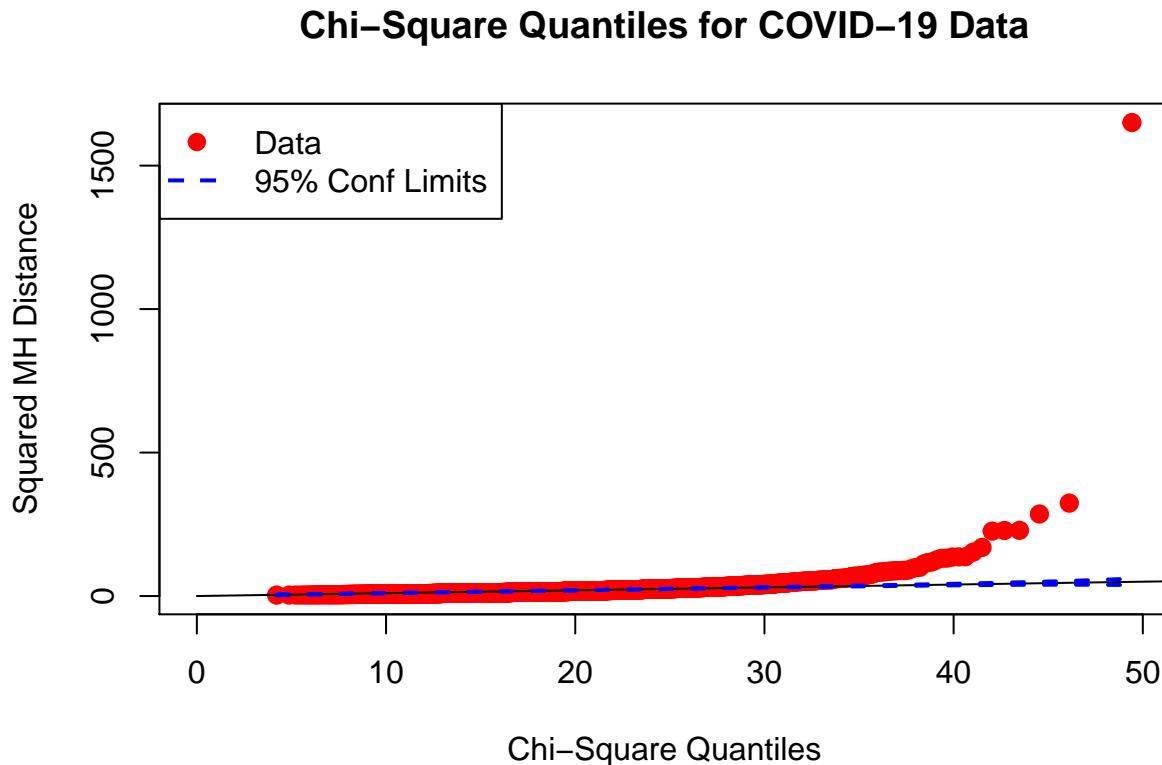
db_trans <- subset(db, select=-trans)

# refresh names
names_trans <- names(db_trans)

for(i in 1:length(names_trans)){
  qqPlot(db_trans[[names_trans[i]]])
}
# looking much more linear

```

```
source("http://www.reuningscherer.net/multivariate/R/CSQPlot.r.txt")
CSQPlot(db_trans[, -1], label="COVID-19 Data")
```



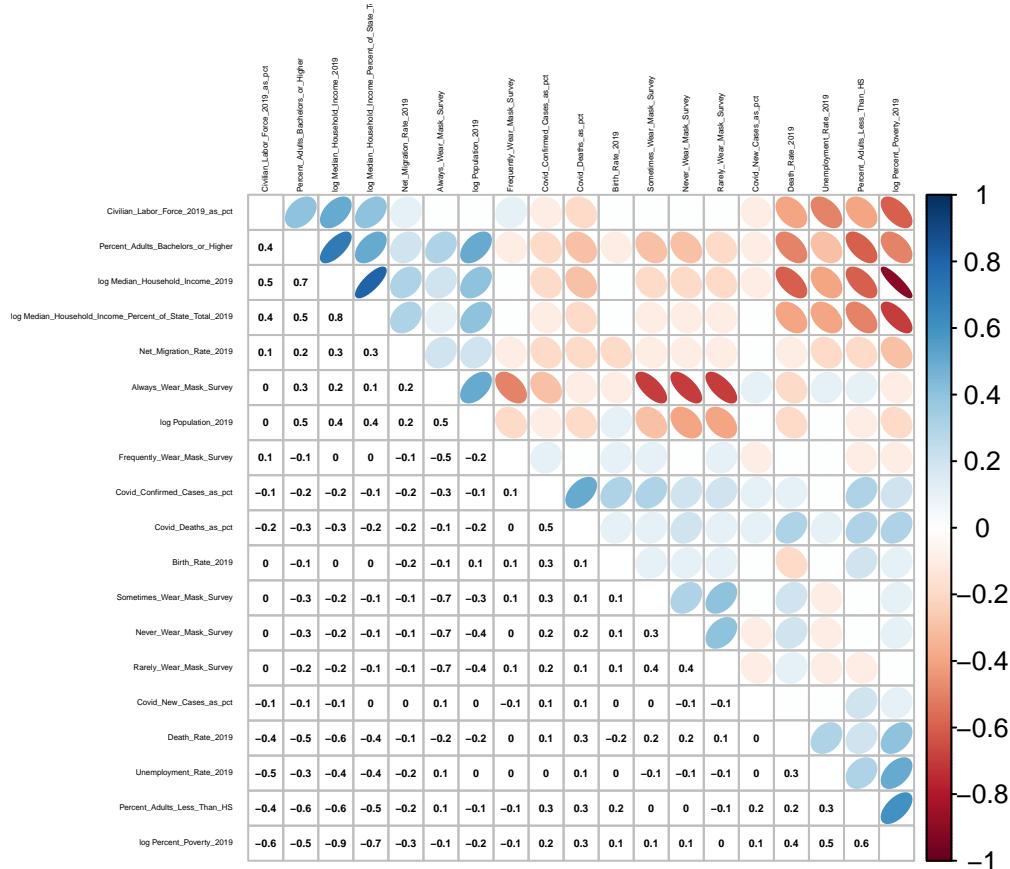
After visually inspecting a boxplot of each of the variables, it was determined that there were 4 variables with a skewed distribution, all of which had many outliers above the boxplot. The median household income, median household income as a percent of the state total, poverty percentage, and population estimate were all transformed using a natural log. After the transformation, the boxplots appeared more in line with the other variables, and the normal quantiles plots were much more linear, with only slight deviations in end behavior.

The chi-square plot deviates from linearity at high chi-square quantiles, indicating that the data does not follow a multivariate normal distribution. This means we can still use PCA to analyze but cannot use parallel analysis to determine which components to retain.

2: Correlation Matrix

```
# round(cor(db_trans[, -1]), 2)

corrplot.mixed(round(cor(db_trans[, -1]), 1), lower.col = "black", upper = "ellipse", tl.col = "black",
```



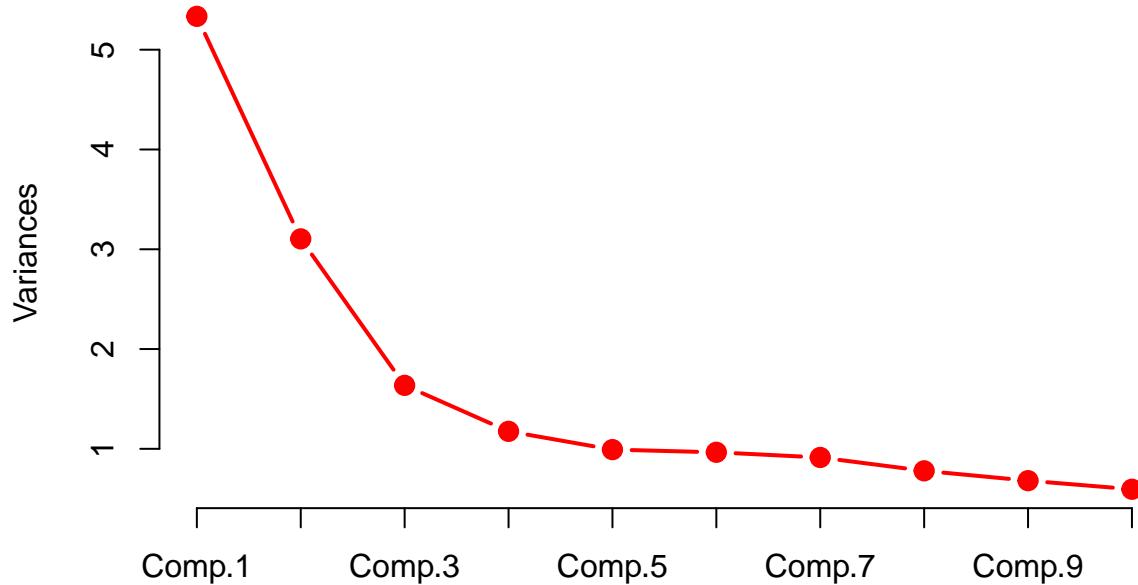
PCA should work well because there are many variables highly correlated with other variables. For instance, always mask and rarely mask have a correlation of -0.73, percentage of adults with less than a high school degree and percentage of adults with a bachelor's degree or higher is -0.60, percentage of adults with a bachelor's degree or higher and the log of the median household income is 0.69. There appear to be underlying trends about the counties (about beliefs about COVID-19, about wealth/education, etc) that could be summarized in linear combinations of the 19 metric variables we have currently.

3: PCA Using Correlation Matrix

```
pc1 <- princomp(db_trans[, -1], cor=TRUE) # run PCA on correlation matrix

screeplot(pc1, type="lines", col="red", lwd=2, pch=19, cex=1.2, main="Scree Plot of Transformed COVID-19 Data")
```

Scree Plot of Transformed COVID-19 Data



```
print(summary(pc1), digits = 2, loadings=pc1$loadings, cutoff=0) # results
```

```
## Importance of components:  
##           Comp.1    Comp.2    Comp.3    Comp.4    Comp.5  
## Standard deviation 2.3095987 1.7618506 1.27912360 1.08406321 0.9960585  
## Proportion of Variance 0.2807498 0.1633746 0.08611354 0.06185226 0.0522175  
## Cumulative Proportion 0.2807498 0.4441244 0.53023794 0.59209020 0.6443077  
##           Comp.6    Comp.7    Comp.8    Comp.9    Comp.10  
## Standard deviation 0.98248645 0.95570754 0.88290685 0.82528062 0.77136488  
## Proportion of Variance 0.05080419 0.04807247 0.04102761 0.03584674 0.03131599  
## Cumulative Proportion 0.69511189 0.74318436 0.78421197 0.82005871 0.85137470  
##           Comp.11   Comp.12   Comp.13   Comp.14   Comp.15  
## Standard deviation 0.74538883 0.7319096 0.69800172 0.64127158 0.59025051  
## Proportion of Variance 0.02924234 0.0281943 0.02564244 0.02164364 0.01833661  
## Cumulative Proportion 0.88061704 0.9088113 0.93445378 0.95609743 0.97443404  
##           Comp.16   Comp.17   Comp.18   Comp.19  
## Standard deviation 0.48274710 0.43616962 0.249906473 3.355940e-03  
## Proportion of Variance 0.01226551 0.01001284 0.003287013 5.927545e-07  
## Cumulative Proportion 0.98669956 0.99671239 0.999999407 1.000000e+00
```

```
## Warning in if (loadings) {: the condition has length > 1 and only the first  
## element will be used
```

```
##  
## Loadings:
```

	Comp.1	Comp.2	Comp.3
##			
## Never_Wear_Mask_Survey	0.17	0.33	0.02
## Rarely_Wear_Mask_Survey	0.13	0.38	0.07
## Sometimes_Wear_Mask_Survey	0.16	0.32	0.01
## Frequently_Wear_Mask_Survey	0.03	0.24	0.09
## Always_Wear_Mask_Survey	-0.19	-0.49	-0.07
## Unemployment_Rate_2019	0.20	-0.26	0.11
## Percent_Adults_Less_Than_HS	0.27	-0.22	-0.24
## Percent_Adults_Bachelors_or_Higher	-0.35	-0.01	-0.05
## Net_Migration_Rate_2019	-0.16	-0.04	0.17
## Death_Rate_2019	0.25	-0.06	0.26
## Birth_Rate_2019	0.05	0.08	-0.52
## Covid_Confirmed_Cases_as_pct	0.17	0.16	-0.51
## Covid_Deaths_as_pct	0.21	0.03	-0.36
## Covid_New_Cases_as_pct	0.04	-0.10	-0.20
## Civilian_Labor_Force_2019_as_pct	-0.24	0.24	-0.03
## log_Median_Household_Income_2019	-0.39	0.11	-0.11
## log_Median_Household_Income_Percent_of_State_Total_2019	-0.33	0.15	-0.17
## log_Percent_Poverty_2019	0.35	-0.21	-0.01
## log_Population_2019	-0.22	-0.22	-0.25
##	Comp.4	Comp.5	Comp.6
## Never_Wear_Mask_Survey	0.07	0.07	0.28
## Rarely_Wear_Mask_Survey	0.00	-0.04	0.22
## Sometimes_Wear_Mask_Survey	0.15	-0.06	0.25
## Frequently_Wear_Mask_Survey	-0.45	-0.30	-0.56
## Always_Wear_Mask_Survey	0.10	0.13	-0.05
## Unemployment_Rate_2019	-0.32	-0.03	0.09
## Percent_Adults_Less_Than_HS	0.00	-0.22	0.04
## Percent_Adults_Bachelors_or_Higher	-0.04	0.18	-0.02
## Net_Migration_Rate_2019	0.28	-0.48	0.30
## Death_Rate_2019	0.26	0.21	0.01
## Birth_Rate_2019	-0.43	-0.15	0.25
## Covid_Confirmed_Cases_as_pct	0.14	0.11	-0.07
## Covid_Deaths_as_pct	0.31	0.33	-0.29
## Covid_New_Cases_as_pct	0.42	-0.61	-0.25
## Civilian_Labor_Force_2019_as_pct	0.12	0.05	-0.23
## log_Median_Household_Income_2019	0.01	0.01	0.00
## log_Median_Household_Income_Percent_of_State_Total_2019	0.04	-0.05	0.10
## log_Percent_Poverty_2019	-0.09	-0.04	0.04
## log_Population_2019	-0.08	0.03	0.34
##	Comp.7	Comp.8	Comp.9
## Never_Wear_Mask_Survey	0.18	0.05	0.51
## Rarely_Wear_Mask_Survey	0.07	0.16	0.24
## Sometimes_Wear_Mask_Survey	-0.13	0.13	-0.76
## Frequently_Wear_Mask_Survey	-0.38	-0.11	0.08
## Always_Wear_Mask_Survey	0.12	-0.08	-0.02
## Unemployment_Rate_2019	-0.21	0.26	0.00
## Percent_Adults_Less_Than_HS	0.18	-0.27	-0.10
## Percent_Adults_Bachelors_or_Higher	-0.03	0.21	0.11
## Net_Migration_Rate_2019	-0.30	-0.56	0.16
## Death_Rate_2019	-0.40	0.12	0.10
## Birth_Rate_2019	0.12	-0.04	0.04
## Covid_Confirmed_Cases_as_pct	-0.25	-0.12	0.00
## Covid_Deaths_as_pct	-0.23	-0.14	0.13

## Covid_New_Cases_as_pct	0.11	0.54	0.11
## Civilian_Labor_Force_2019_as_pct	0.36	-0.24	-0.10
## log_Median_Household_Income_2019	-0.07	0.07	-0.02
## log_Median_Household_Income_Percent_of_State_Total_2019	-0.25	0.10	-0.02
## log_Percent_Poverty_2019	0.08	-0.05	0.03
## log_Population_2019	-0.35	0.13	0.07
##	Comp.10	Comp.11	Comp.12
## Never_Wear_Mask_Survey	0.09	0.54	0.26
## Rarely_Wear_Mask_Survey	0.04	-0.72	-0.26
## Sometimes_Wear_Mask_Survey	-0.08	0.13	0.13
## Frequently_Wear_Mask_Survey	-0.14	0.07	0.08
## Always_Wear_Mask_Survey	0.04	-0.02	-0.09
## Unemployment_Rate_2019	0.48	0.12	-0.13
## Percent_Adults_Less_Than_HS	0.19	-0.01	-0.05
## Percent_Adults_Bachelors_or_Higher	-0.22	-0.12	0.38
## Net_Migration_Rate_2019	0.01	-0.04	0.04
## Death_Rate_2019	-0.32	0.21	-0.47
## Birth_Rate_2019	-0.25	0.16	-0.39
## Covid_Confirmed_Cases_as_pct	0.10	-0.14	0.28
## Covid_Deaths_as_pct	0.12	-0.02	-0.12
## Covid_New_Cases_as_pct	-0.08	0.06	-0.02
## Civilian_Labor_Force_2019_as_pct	-0.22	0.16	-0.30
## log_Median_Household_Income_2019	0.25	0.07	-0.13
## log_Median_Household_Income_Percent_of_State_Total_2019	0.34	0.09	-0.13
## log_Percent_Poverty_2019	-0.27	-0.10	0.27
## log_Population_2019	-0.40	-0.02	0.04
##	Comp.13	Comp.14	Comp.15
## Never_Wear_Mask_Survey	0.05	0.06	0.09
## Rarely_Wear_Mask_Survey	-0.02	0.09	0.11
## Sometimes_Wear_Mask_Survey	-0.13	-0.09	0.12
## Frequently_Wear_Mask_Survey	0.10	0.02	0.11
## Always_Wear_Mask_Survey	0.00	-0.03	-0.17
## Unemployment_Rate_2019	-0.54	0.30	0.06
## Percent_Adults_Less_Than_HS	0.38	0.17	0.35
## Percent_Adults_Bachelors_or_Higher	-0.28	-0.13	0.07
## Net_Migration_Rate_2019	-0.31	-0.10	-0.06
## Death_Rate_2019	0.15	0.08	-0.20
## Birth_Rate_2019	-0.17	-0.33	-0.22
## Covid_Confirmed_Cases_as_pct	-0.01	0.41	-0.54
## Covid_Deaths_as_pct	-0.28	-0.36	0.45
## Covid_New_Cases_as_pct	-0.08	0.00	-0.03
## Civilian_Labor_Force_2019_as_pct	-0.35	0.55	0.16
## log_Median_Household_Income_2019	0.09	-0.04	-0.02
## log_Median_Household_Income_Percent_of_State_Total_2019	0.26	0.06	0.14
## log_Percent_Poverty_2019	-0.10	0.08	0.13
## log_Population_2019	0.12	0.33	0.38
##	Comp.16	Comp.17	Comp.18
## Never_Wear_Mask_Survey	0.04	0.04	0.01
## Rarely_Wear_Mask_Survey	-0.02	0.04	0.03
## Sometimes_Wear_Mask_Survey	0.00	0.07	0.02
## Frequently_Wear_Mask_Survey	0.01	0.05	0.01
## Always_Wear_Mask_Survey	-0.01	-0.08	-0.03
## Unemployment_Rate_2019	-0.03	0.09	-0.02
## Percent_Adults_Less_Than_HS	-0.28	0.49	-0.03

```

## Percent_Adults_Bachelors_or_Higher           -0.52   0.43  -0.10
## Net_Migration_Rate_2019                     -0.05   0.04   0.01
## Death_Rate_2019                            -0.29   0.20   0.09
## Birth_Rate_2019                            -0.10   0.01  -0.03
## Covid_Confirmed_Cases_as_pct                0.01    0.05  -0.01
## Covid_Deaths_as_pct                         0.10   -0.06   0.01
## Covid_New_Cases_as_pct                      0.04   -0.01  -0.01
## Civilian_Labor_Force_2019_as_pct            -0.09   -0.05   0.00
## log_Median_Household_Income_2019            0.11    0.26   0.80
## log_Median_Household_Income_Percent_of_State_Total_2019 -0.51   -0.49  -0.16
## log_Percent_Poverty_2019                    -0.34   -0.44   0.56
## log_Population_2019                          0.37   -0.02  -0.05
##                                         Comp.19
## Never_Wear_Mask_Survey                      0.30
## Rarely_Wear_Mask_Survey                     0.29
## Sometimes_Wear_Mask_Survey                  0.30
## Frequently_Wear_Mask_Survey                 0.33
## Always_Wear_Mask_Survey                     0.79
## Unemployment_Rate_2019                      0.00
## Percent_Adults_Less_Than_HS                 0.00
## Percent_Adults_Bachelors_or_Higher          0.00
## Net_Migration_Rate_2019                      0.00
## Death_Rate_2019                            0.00
## Birth_Rate_2019                            0.00
## Covid_Confirmed_Cases_as_pct                0.00
## Covid_Deaths_as_pct                         0.00
## Covid_New_Cases_as_pct                      0.00
## Civilian_Labor_Force_2019_as_pct            0.00
## log_Median_Household_Income_2019            0.00
## log_Median_Household_Income_Percent_of_State_Total_2019 0.00
## log_Percent_Poverty_2019                    0.00
## log_Population_2019                          0.00

round(pc1$sdev^2, 2) # eigenvalues

##   Comp.1  Comp.2  Comp.3  Comp.4  Comp.5  Comp.6  Comp.7  Comp.8  Comp.9  Comp.10
##     5.33    3.10    1.64    1.18    0.99    0.97    0.91    0.78    0.68    0.60
##   Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17 Comp.18 Comp.19
##     0.56    0.54    0.49    0.41    0.35    0.23    0.19    0.06    0.00

```

According to the total variance explained method, using a cutoff of 0.8, the first 9 PC's should be used. According to the Eigenvalue > 1 method, the first 4 PC's should be used. According to the scree plot elbow method, the first 3 PC's should be used.

To avoid an unnecessarily complicated model, I will choose to use the scree plot elbow method and use the first 3 PC's.

Note that parallel analysis, though the preferred method, would not have been appropriate here because the data does not follow a multivariate normal distribution.

4: Principle Component Examination

```
print(summary(pc1), digits = 2, loadings=pc1$loadings, cutoff=0) # summary

## Importance of components:
##          Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
## Standard deviation 2.3095987 1.7618506 1.27912360 1.08406321 0.9960585
## Proportion of Variance 0.2807498 0.1633746 0.08611354 0.06185226 0.0522175
## Cumulative Proportion 0.2807498 0.4441244 0.53023794 0.59209020 0.6443077
##          Comp.6    Comp.7    Comp.8    Comp.9    Comp.10
## Standard deviation 0.98248645 0.95570754 0.88290685 0.82528062 0.77136488
## Proportion of Variance 0.05080419 0.04807247 0.04102761 0.03584674 0.03131599
## Cumulative Proportion 0.69511189 0.74318436 0.78421197 0.82005871 0.85137470
##          Comp.11   Comp.12   Comp.13   Comp.14   Comp.15
## Standard deviation 0.74538883 0.7319096 0.69800172 0.64127158 0.59025051
## Proportion of Variance 0.02924234 0.0281943 0.02564244 0.02164364 0.01833661
## Cumulative Proportion 0.88061704 0.9088113 0.93445378 0.95609743 0.97443404
##          Comp.16   Comp.17   Comp.18   Comp.19
## Standard deviation 0.48274710 0.43616962 0.249906473 3.355940e-03
## Proportion of Variance 0.01226551 0.01001284 0.003287013 5.927545e-07
## Cumulative Proportion 0.98669956 0.99671239 0.999999407 1.000000e+00

## Warning in if (loadings) {: the condition has length > 1 and only the first
## element will be used

## Loadings:
##          Never_Wear_Mask_Survey  Rarely_Wear_Mask_Survey  Sometimes_Wear_Mask_Survey
##          Frequently_Wear_Mask_Survey  Always_Wear_Mask_Survey  Unemployment_Rate_2019
##          Percent_Adults_Less_Than_HS  Percent_Adults_Bachelors_or_Higher  Net_Migration_Rate_2019
##          Death_Rate_2019  Birth_Rate_2019  Covid_Confirmed_Cases_as_pct
##          Covid_Deaths_as_pct  Covid_New_Cases_as_pct  Civilian_Labor_Force_2019_as_pct
##          log_Median_Household_Income_2019  log_Median_Household_Income_Percent_of_State_Total_2019
##          log_Percent_Poverty_2019  log_Population_2019
##          Never_Wear_Mask_Survey  Rarely_Wear_Mask_Survey  Sometimes_Wear_Mask_Survey
##          Frequently_Wear_Mask_Survey
##          Comp.1    Comp.2    Comp.3
##          0.17    0.33    0.02
##          0.13    0.38    0.07
##          0.16    0.32    0.01
##          0.03    0.24    0.09
##          -0.19   -0.49   -0.07
##          0.20    -0.26   0.11
##          0.27    -0.22   -0.24
##          -0.35   -0.01   -0.05
##          -0.16   -0.04   0.17
##          0.25    -0.06   0.26
##          0.05    0.08   -0.52
##          0.17    0.16   -0.51
##          0.21    0.03   -0.36
##          0.04   -0.10   -0.20
##          -0.24   0.24   -0.03
##          -0.39   0.11   -0.11
##          -0.33   0.15   -0.17
##          0.35   -0.21   -0.01
##          -0.22   -0.22   -0.25
##          Comp.4    Comp.5    Comp.6
##          0.07    0.07    0.28
##          0.00   -0.04    0.22
##          0.15   -0.06    0.25
##          -0.45   -0.30   -0.56
```

## Always_Wear_Mask_Survey	0.10	0.13	-0.05
## Unemployment_Rate_2019	-0.32	-0.03	0.09
## Percent_Adults_Less_Than_HS	0.00	-0.22	0.04
## Percent_Adults_Bachelors_or_Higher	-0.04	0.18	-0.02
## Net_Migration_Rate_2019	0.28	-0.48	0.30
## Death_Rate_2019	0.26	0.21	0.01
## Birth_Rate_2019	-0.43	-0.15	0.25
## Covid_Confirmed_Cases_as_pct	0.14	0.11	-0.07
## Covid_Deaths_as_pct	0.31	0.33	-0.29
## Covid_New_Cases_as_pct	0.42	-0.61	-0.25
## Civilian_Labor_Force_2019_as_pct	0.12	0.05	-0.23
## log_Median_Household_Income_2019	0.01	0.01	0.00
## log_Median_Household_Income_Percent_of_State_Total_2019	0.04	-0.05	0.10
## log_Percent_Poverty_2019	-0.09	-0.04	0.04
## log_Population_2019	-0.08	0.03	0.34
##	Comp.7	Comp.8	Comp.9
## Never_Wear_Mask_Survey	0.18	0.05	0.51
## Rarely_Wear_Mask_Survey	0.07	0.16	0.24
## Sometimes_Wear_Mask_Survey	-0.13	0.13	-0.76
## Frequently_Wear_Mask_Survey	-0.38	-0.11	0.08
## Always_Wear_Mask_Survey	0.12	-0.08	-0.02
## Unemployment_Rate_2019	-0.21	0.26	0.00
## Percent_Adults_Less_Than_HS	0.18	-0.27	-0.10
## Percent_Adults_Bachelors_or_Higher	-0.03	0.21	0.11
## Net_Migration_Rate_2019	-0.30	-0.56	0.16
## Death_Rate_2019	-0.40	0.12	0.10
## Birth_Rate_2019	0.12	-0.04	0.04
## Covid_Confirmed_Cases_as_pct	-0.25	-0.12	0.00
## Covid_Deaths_as_pct	-0.23	-0.14	0.13
## Covid_New_Cases_as_pct	0.11	0.54	0.11
## Civilian_Labor_Force_2019_as_pct	0.36	-0.24	-0.10
## log_Median_Household_Income_2019	-0.07	0.07	-0.02
## log_Median_Household_Income_Percent_of_State_Total_2019	-0.25	0.10	-0.02
## log_Percent_Poverty_2019	0.08	-0.05	0.03
## log_Population_2019	-0.35	0.13	0.07
##	Comp.10	Comp.11	Comp.12
## Never_Wear_Mask_Survey	0.09	0.54	0.26
## Rarely_Wear_Mask_Survey	0.04	-0.72	-0.26
## Sometimes_Wear_Mask_Survey	-0.08	0.13	0.13
## Frequently_Wear_Mask_Survey	-0.14	0.07	0.08
## Always_Wear_Mask_Survey	0.04	-0.02	-0.09
## Unemployment_Rate_2019	0.48	0.12	-0.13
## Percent_Adults_Less_Than_HS	0.19	-0.01	-0.05
## Percent_Adults_Bachelors_or_Higher	-0.22	-0.12	0.38
## Net_Migration_Rate_2019	0.01	-0.04	0.04
## Death_Rate_2019	-0.32	0.21	-0.47
## Birth_Rate_2019	-0.25	0.16	-0.39
## Covid_Confirmed_Cases_as_pct	0.10	-0.14	0.28
## Covid_Deaths_as_pct	0.12	-0.02	-0.12
## Covid_New_Cases_as_pct	-0.08	0.06	-0.02
## Civilian_Labor_Force_2019_as_pct	-0.22	0.16	-0.30
## log_Median_Household_Income_2019	0.25	0.07	-0.13
## log_Median_Household_Income_Percent_of_State_Total_2019	0.34	0.09	-0.13
## log_Percent_Poverty_2019	-0.27	-0.10	0.27

## log Population_2019	-0.40	-0.02	0.04
##	Comp.13	Comp.14	Comp.15
## Never_Wear_Mask_Survey	0.05	0.06	0.09
## Rarely_Wear_Mask_Survey	-0.02	0.09	0.11
## Sometimes_Wear_Mask_Survey	-0.13	-0.09	0.12
## Frequently_Wear_Mask_Survey	0.10	0.02	0.11
## Always_Wear_Mask_Survey	0.00	-0.03	-0.17
## Unemployment_Rate_2019	-0.54	0.30	0.06
## Percent_Adults_Less_Than_HS	0.38	0.17	0.35
## Percent_Adults_Bachelors_or_Higher	-0.28	-0.13	0.07
## Net_Migration_Rate_2019	-0.31	-0.10	-0.06
## Death_Rate_2019	0.15	0.08	-0.20
## Birth_Rate_2019	-0.17	-0.33	-0.22
## Covid_Confirmed_Cases_as_pct	-0.01	0.41	-0.54
## Covid_Deaths_as_pct	-0.28	-0.36	0.45
## Covid_New_Cases_as_pct	-0.08	0.00	-0.03
## Civilian_Labor_Force_2019_as_pct	-0.35	0.55	0.16
## log Median_Household_Income_2019	0.09	-0.04	-0.02
## log Median_Household_Income_Percent_of_State_Total_2019	0.26	0.06	0.14
## log Percent_Poverty_2019	-0.10	0.08	0.13
## log Population_2019	0.12	0.33	0.38
##	Comp.16	Comp.17	Comp.18
## Never_Wear_Mask_Survey	0.04	0.04	0.01
## Rarely_Wear_Mask_Survey	-0.02	0.04	0.03
## Sometimes_Wear_Mask_Survey	0.00	0.07	0.02
## Frequently_Wear_Mask_Survey	0.01	0.05	0.01
## Always_Wear_Mask_Survey	-0.01	-0.08	-0.03
## Unemployment_Rate_2019	-0.03	0.09	-0.02
## Percent_Adults_Less_Than_HS	-0.28	0.49	-0.03
## Percent_Adults_Bachelors_or_Higher	-0.52	0.43	-0.10
## Net_Migration_Rate_2019	-0.05	0.04	0.01
## Death_Rate_2019	-0.29	0.20	0.09
## Birth_Rate_2019	-0.10	0.01	-0.03
## Covid_Confirmed_Cases_as_pct	0.01	0.05	-0.01
## Covid_Deaths_as_pct	0.10	-0.06	0.01
## Covid_New_Cases_as_pct	0.04	-0.01	-0.01
## Civilian_Labor_Force_2019_as_pct	-0.09	-0.05	0.00
## log Median_Household_Income_2019	0.11	0.26	0.80
## log Median_Household_Income_Percent_of_State_Total_2019	-0.51	-0.49	-0.16
## log Percent_Poverty_2019	-0.34	-0.44	0.56
## log Population_2019	0.37	-0.02	-0.05
##	Comp.19		
## Never_Wear_Mask_Survey	0.30		
## Rarely_Wear_Mask_Survey	0.29		
## Sometimes_Wear_Mask_Survey	0.30		
## Frequently_Wear_Mask_Survey	0.33		
## Always_Wear_Mask_Survey	0.79		
## Unemployment_Rate_2019	0.00		
## Percent_Adults_Less_Than_HS	0.00		
## Percent_Adults_Bachelors_or_Higher	0.00		
## Net_Migration_Rate_2019	0.00		
## Death_Rate_2019	0.00		
## Birth_Rate_2019	0.00		
## Covid_Confirmed_Cases_as_pct	0.00		

```

## Covid_Deaths_as_pct                      0.00
## Covid_New_Cases_as_pct                   0.00
## Civilian_Labor_Force_2019_as_pct        0.00
## log_Median_Household_Income_2019        0.00
## log_Median_Household_Income_Percent_of_State_Total_2019 0.00
## log_Percent_Poverty_2019                 0.00
## log_Population_2019                      0.00

```

Looking at PC1: This principle component seems to be related to wealth and education. It combines log median household income (-0.39) with the log of the percent below the poverty line (0.35) and the percent of adults with a bachelor's degree or higher (-0.35).

Looking at PC2: This principle component seems to be a measure of masking behaviors. It combines the percentage of those who say they always mask (-0.49), rarely mask mask (0.38), and never mask (0.33).

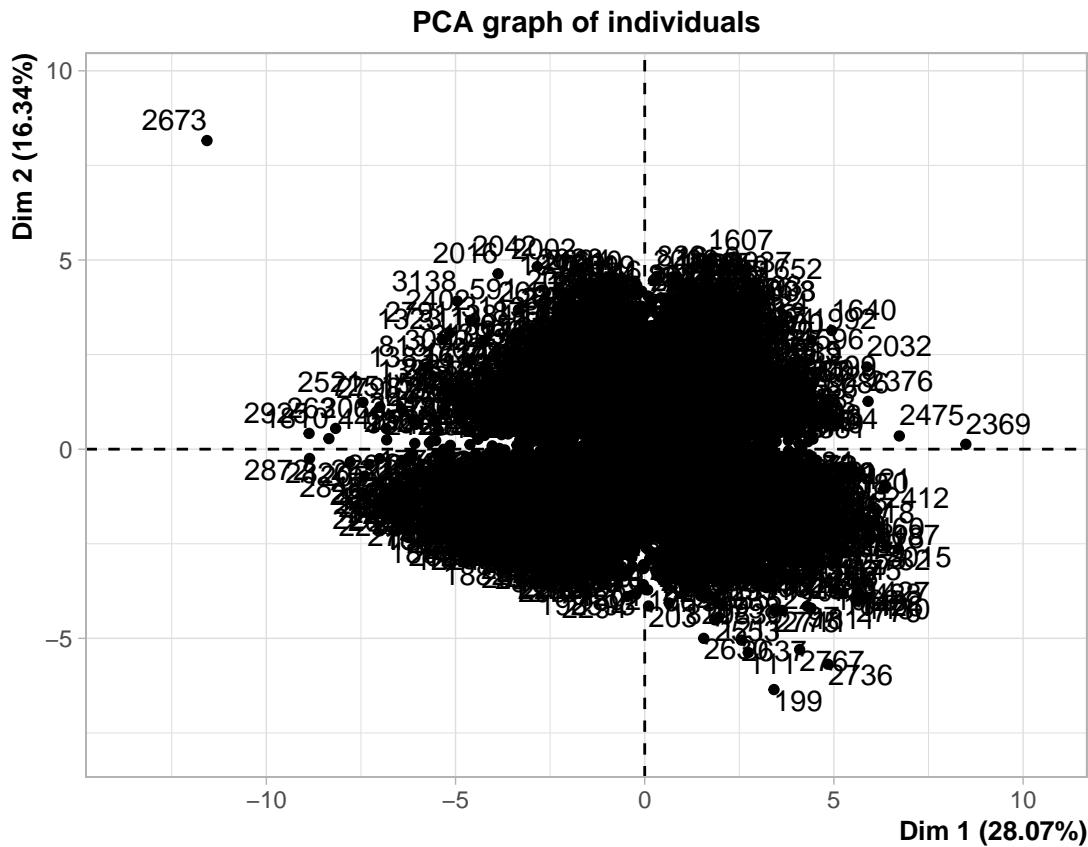
Looking at PC3: This principle component seems to be a measure of population as it relates to confirmed cases. It combines the birth rate (-0.52) with the cumulative percentage of population with confirmed cases (-0.51) and the cumulative percentage of population that has died of COVID-19 (-0.36).

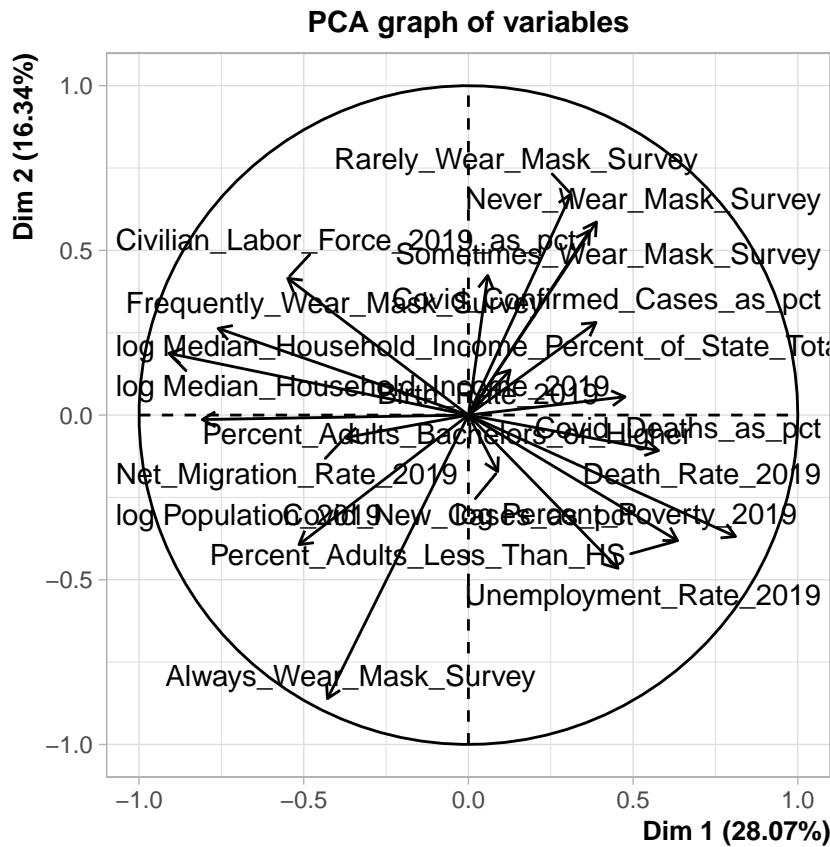
5: Score Plot

```

library(FactoMineR)
par(cex=.4)
pc2 <- PCA(db_trans[, -1])

```



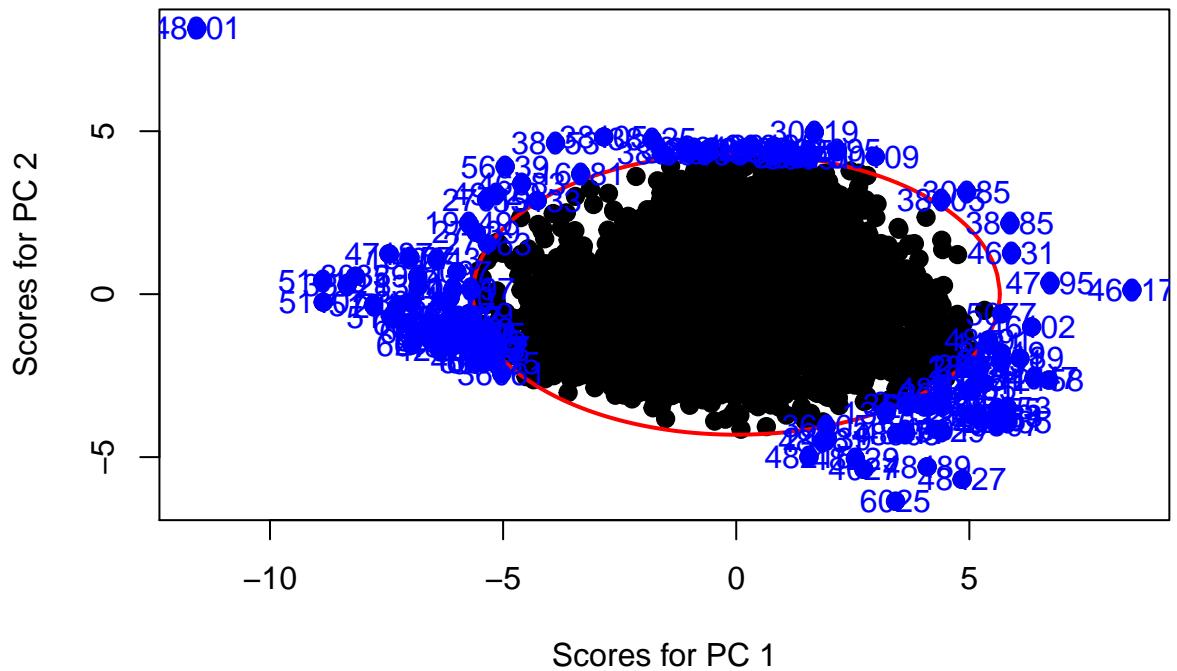


There are no noticeable trends in the score plot, though observation 2673 is an outlier (FIPS 48301 in Loving, Texas, which has a very small population of 169, the lowest population in the entire dataset).

In the biplot, quadrant 1 is associated with poor adherence to masking and high percentages of confirmed cases; quadrant 2 is associated with frequent masking, high income, and high education; quadrant 3 is associated with population and new cases as of February 18, 2021; and quadrant 4 is associated with the unemployment rate, poverty rate, and death rate. These groupings make sense because PC 1 is related to wealth and education and PC 2 is related to masking behaviors.

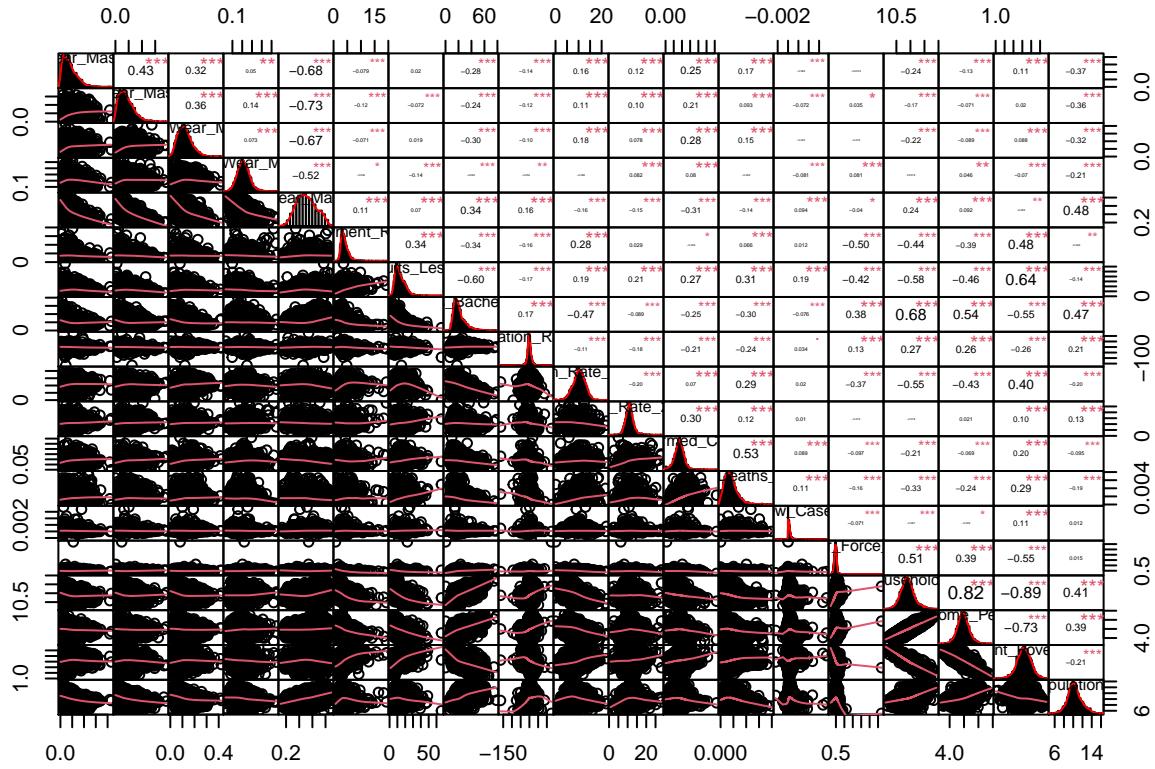
```
source("http://reuningscherer.net/multivariate/r/ciscoreplot.R.txt")
ciscoreplot(pc1, c(1,2), db_trans[, 1])
```

PC Score Plot with 95% CI Ellipse



6: Summary

```
chart.Correlation(db_trans[, -1], histogram=TRUE, pch=19)
```



This data set contains COVID-19 infection and death statistics from U.S. counties, combined with economic, education, and population data and also survey responses about mask-wearing frequencies. It has 3141 complete observations on 19 metric variables; since the number of observations is over 150 times the number of variables, we have more than enough observations given our dimensions.

Most of the relationships appear to be linear, and, after log transforming a few variables, the variables follow approximately normal distributions. However, the data does not completely follow a multivariate normal distribution, according to the chi-square plot, which deviates from linearity at the ends.

Using PCA, we can reduce these 19 variables to 3 composite variables that are related to wealth and education, attitudes about masking, and population. These 3 PC's can account for 53% of the total variability, which is moderately effective.

As noted in the score plot, there is a multivariate outlier, with county 48301 (in Loving Texas) having the lowest scores on dimension 1 and dimension 2, indicating high education and high adherence to masking. This is likely connected to the fact that this county has a very small population of 169, the lowest population in the entire dataset, so any sampling bias could greatly skew these estimates based on such a small sample size.