

S&DS 363 Homework 7

Evan Collins, Kelly Farley, Ken Stier

28 April 2021

```
# clear env
rm(list=ls())

# load packages
library(corrplot)
library(tidyverse)
library(rela)
library(psych)
library(dplyr)
```

Contributors

Evan Collins (evan.collins@yale.edu)

Kelly Farley (kelly.farley@yale.edu)

Ken Stier (ken.stier@yale.edu)

The Dataset

Raw 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 and 6 categorical variables. 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. We examine 19 continuous variables, with 3 of them (l)median household income, percent poverty, and percentage of confirmed COVID-19 cases) log transformed as has been indicated in previous psets.

```
# Process raw dataset as we have done in preceding psets

raw <- read.csv("https://evancollins.com/covid_and_demographics.csv")

# create categorical variables: rural-urban code (3 levels), region (4 variables)

# log transformations of our continuous variables
raw$logMedian_Household_Income_2019 <- log(raw$Median_Household_Income_2019 + 0.0001)
raw$logPercent_Poverty_2019 <- log(raw$Percent_Poverty_2019 + 0.0001)
raw$logCovid_Confirmed_Cases_as_pct <- log(raw$Covid_Confirmed_Cases_as_pct + 0.0001)

# only want continuous variables
dat <- raw[, c(5:28)]

# remove untransformed variables and integers
```

```
dat <- dat[, -c(18, 7, 9, 16, 17)]
```

1) Look through indicators (questions). Think about which indicators might be related through latent factors. (nothing to turn in here)

2) Compute the correlation matrix between all indicators (you may want to do this in batches). Comment on relationships you do/do not observe.

```
cor(dat)
```

```
##                                     Never_Wear_Mask_Survey
## Never_Wear_Mask_Survey                1.000000000
## Rarely_Wear_Mask_Survey                0.427676105
## Sometimes_Wear_Mask_Survey             0.322745764
## Frequently_Wear_Mask_Survey             0.049828706
## Always_Wear_Mask_Survey                -0.684577984
## Unemployment_Rate_2019                 -0.079133639
## Median_Household_Income_Percent_of_State_Total_2019 -0.154577804
## Percent_Adults_Less_Than_HS             0.020277222
## Percent_Adults_Bachelors_or_Higher     -0.280334539
## Population_2019                       -0.200463449
## Net_Migration_Rate_2019                -0.139883470
## Death_Rate_2019                       0.160090810
## Birth_Rate_2019                       0.124217737
## Covid_Deaths_as_pct                   0.165192994
## Covid_New_Cases_as_pct                 -0.064959691
## Civilian_Labor_Force_2019_as_pct       -0.001773504
## logMedian_Household_Income_2019        -0.240717720
## logPercent_Poverty_2019                0.113300309
## logCovid_Confirmed_Cases_as_pct        0.120424610
##                                     Rarely_Wear_Mask_Survey
## Never_Wear_Mask_Survey                0.42767611
## Rarely_Wear_Mask_Survey                1.00000000
## Sometimes_Wear_Mask_Survey             0.36412718
## Frequently_Wear_Mask_Survey             0.14352041
## Always_Wear_Mask_Survey                -0.72807963
## Unemployment_Rate_2019                 -0.11585788
## Median_Household_Income_Percent_of_State_Total_2019 -0.09381128
## Percent_Adults_Less_Than_HS             -0.07195757
## Percent_Adults_Bachelors_or_Higher     -0.23895817
## Population_2019                       -0.21593322
## Net_Migration_Rate_2019                -0.11513583
## Death_Rate_2019                       0.11102699
## Birth_Rate_2019                       0.10372687
## Covid_Deaths_as_pct                   0.09335193
## Covid_New_Cases_as_pct                 -0.07158883
## Civilian_Labor_Force_2019_as_pct       0.03528312
## logMedian_Household_Income_2019        -0.17361297
```

## logPercent_Poverty_2019	0.02027919
## logCovid_Confirmed_Cases_as_pct	0.11019918
##	Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey	0.32274576
## Rarely_Wear_Mask_Survey	0.36412718
## Sometimes_Wear_Mask_Survey	1.00000000
## Frequently_Wear_Mask_Survey	0.07300172
## Always_Wear_Mask_Survey	-0.66869703
## Unemployment_Rate_2019	-0.07121451
## Median_Household_Income_Percent_of_State_Total_2019	-0.11003094
## Percent_Adults_Less_Than_HS	0.01870091
## Percent_Adults_Bachelors_or_Higher	-0.30342656
## Population_2019	-0.21082665
## Net_Migration_Rate_2019	-0.10426920
## Death_Rate_2019	0.18079613
## Birth_Rate_2019	0.07824170
## Covid_Deaths_as_pct	0.14562233
## Covid_New_Cases_as_pct	-0.02303206
## Civilian_Labor_Force_2019_as_pct	-0.01621358
## logMedian_Household_Income_2019	-0.22306024
## logPercent_Poverty_2019	0.08775493
## logCovid_Confirmed_Cases_as_pct	0.16479929
##	Frequently_Wear_Mask_Survey
## Never_Wear_Mask_Survey	0.049828706
## Rarely_Wear_Mask_Survey	0.143520412
## Sometimes_Wear_Mask_Survey	0.073001721
## Frequently_Wear_Mask_Survey	1.000000000
## Always_Wear_Mask_Survey	-0.517209567
## Unemployment_Rate_2019	-0.036034054
## Median_Household_Income_Percent_of_State_Total_2019	0.034642515
## Percent_Adults_Less_Than_HS	-0.140631442
## Percent_Adults_Bachelors_or_Higher	-0.063220967
## Population_2019	-0.152509645
## Net_Migration_Rate_2019	-0.053600134
## Death_Rate_2019	-0.028181033
## Birth_Rate_2019	0.081951484
## Covid_Deaths_as_pct	-0.027613685
## Covid_New_Cases_as_pct	-0.081472062
## Civilian_Labor_Force_2019_as_pct	0.081487685
## logMedian_Household_Income_2019	0.001936773
## logPercent_Poverty_2019	-0.070079214
## logCovid_Confirmed_Cases_as_pct	-0.021474934
##	Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey	-0.68457798
## Rarely_Wear_Mask_Survey	-0.72807963
## Sometimes_Wear_Mask_Survey	-0.66869703
## Frequently_Wear_Mask_Survey	-0.51720957
## Always_Wear_Mask_Survey	1.00000000
## Unemployment_Rate_2019	0.11494288
## Median_Household_Income_Percent_of_State_Total_2019	0.12112957
## Percent_Adults_Less_Than_HS	0.07001880
## Percent_Adults_Bachelors_or_Higher	0.33704308
## Population_2019	0.29991380
## Net_Migration_Rate_2019	0.15800706

## Death_Rate_2019	-0.15909794
## Birth_Rate_2019	-0.14992133
## Covid_Deaths_as_pct	-0.14157596
## Covid_New_Cases_as_pct	0.09393703
## Civilian_Labor_Force_2019_as_pct	-0.03997857
## logMedian_Household_Income_2019	0.24010991
## logPercent_Poverty_2019	-0.05519402
## logCovid_Confirmed_Cases_as_pct	-0.14027170
##	Unemployment_Rate_2019
## Never_Wear_Mask_Survey	-0.07913364
## Rarely_Wear_Mask_Survey	-0.11585788
## Sometimes_Wear_Mask_Survey	-0.07121451
## Frequently_Wear_Mask_Survey	-0.03603405
## Always_Wear_Mask_Survey	0.11494288
## Unemployment_Rate_2019	1.00000000
## Median_Household_Income_Percent_of_State_Total_2019	-0.36811072
## Percent_Adults_Less_Than_HS	0.33913887
## Percent_Adults_Bachelors_or_Higher	-0.33740011
## Population_2019	-0.05697060
## Net_Migration_Rate_2019	-0.16283518
## Death_Rate_2019	0.27719318
## Birth_Rate_2019	0.02863821
## Covid_Deaths_as_pct	0.06632755
## Covid_New_Cases_as_pct	0.01246554
## Civilian_Labor_Force_2019_as_pct	-0.50174142
## logMedian_Household_Income_2019	-0.43808371
## logPercent_Poverty_2019	0.48249406
## logCovid_Confirmed_Cases_as_pct	-0.03076392
##	Median_Household_Income_Percent_of_State_Total_2019
## Never_Wear_Mask_Survey	-0.15457741
## Rarely_Wear_Mask_Survey	-0.09381111
## Sometimes_Wear_Mask_Survey	-0.11003091
## Frequently_Wear_Mask_Survey	0.03464221
## Always_Wear_Mask_Survey	0.12112991
## Unemployment_Rate_2019	-0.36811072
## Median_Household_Income_Percent_of_State_Total_2019	1.00000000
## Percent_Adults_Less_Than_HS	-0.43942221
## Percent_Adults_Bachelors_or_Higher	0.57322111
## Population_2019	0.20907881
## Net_Migration_Rate_2019	0.26481741
## Death_Rate_2019	-0.44257111
## Birth_Rate_2019	0.01921211
## Covid_Deaths_as_pct	-0.23747211
## Covid_New_Cases_as_pct	-0.02704511
## Civilian_Labor_Force_2019_as_pct	0.37936611
## logMedian_Household_Income_2019	0.81787911
## logPercent_Poverty_2019	-0.72740411
## logCovid_Confirmed_Cases_as_pct	-0.00518311
##	Percent_Adults_Less_Than_HS
## Never_Wear_Mask_Survey	0.02027722
## Rarely_Wear_Mask_Survey	-0.07195757
## Sometimes_Wear_Mask_Survey	0.01870091
## Frequently_Wear_Mask_Survey	-0.14063144
## Always_Wear_Mask_Survey	0.07001880

## Unemployment_Rate_2019	0.33913887
## Median_Household_Income_Percent_of_State_Total_2019	-0.43942284
## Percent_Adults_Less_Than_HS	1.00000000
## Percent_Adults_Bachelors_or_Higher	-0.59775900
## Population_2019	-0.05085429
## Net_Migration_Rate_2019	-0.16831695
## Death_Rate_2019	0.18693289
## Birth_Rate_2019	0.21065691
## Covid_Deaths_as_pct	0.30761178
## Covid_New_Cases_as_pct	0.19292024
## Civilian_Labor_Force_2019_as_pct	-0.42275824
## logMedian_Household_Income_2019	-0.57591742
## logPercent_Poverty_2019	0.63926157
## logCovid_Confirmed_Cases_as_pct	0.20751928
##	Percent_Adults_Bachelors_or_Higher
## Never_Wear_Mask_Survey	-0.28033454
## Rarely_Wear_Mask_Survey	-0.23895817
## Sometimes_Wear_Mask_Survey	-0.30342656
## Frequently_Wear_Mask_Survey	-0.06322097
## Always_Wear_Mask_Survey	0.33704308
## Unemployment_Rate_2019	-0.33740011
## Median_Household_Income_Percent_of_State_Total_2019	0.57322177
## Percent_Adults_Less_Than_HS	-0.59775900
## Percent_Adults_Bachelors_or_Higher	1.00000000
## Population_2019	0.33070934
## Net_Migration_Rate_2019	0.17283606
## Death_Rate_2019	-0.47306789
## Birth_Rate_2019	-0.08904625
## Covid_Deaths_as_pct	-0.30057177
## Covid_New_Cases_as_pct	-0.07649416
## Civilian_Labor_Force_2019_as_pct	0.37697353
## logMedian_Household_Income_2019	0.68475442
## logPercent_Poverty_2019	-0.54889770
## logCovid_Confirmed_Cases_as_pct	-0.16552684
##	Population_2019
## Never_Wear_Mask_Survey	-0.200463449
## Rarely_Wear_Mask_Survey	-0.215933224
## Sometimes_Wear_Mask_Survey	-0.210826654
## Frequently_Wear_Mask_Survey	-0.152509645
## Always_Wear_Mask_Survey	0.299913801
## Unemployment_Rate_2019	-0.056970596
## Median_Household_Income_Percent_of_State_Total_2019	0.209078484
## Percent_Adults_Less_Than_HS	-0.050854288
## Percent_Adults_Bachelors_or_Higher	0.330709335
## Population_2019	1.000000000
## Net_Migration_Rate_2019	0.040484875
## Death_Rate_2019	-0.205792285
## Birth_Rate_2019	0.076483496
## Covid_Deaths_as_pct	-0.059135047
## Covid_New_Cases_as_pct	0.010817944
## Civilian_Labor_Force_2019_as_pct	0.086344982
## logMedian_Household_Income_2019	0.269169601
## logPercent_Poverty_2019	-0.117459035
## logCovid_Confirmed_Cases_as_pct	0.002671028

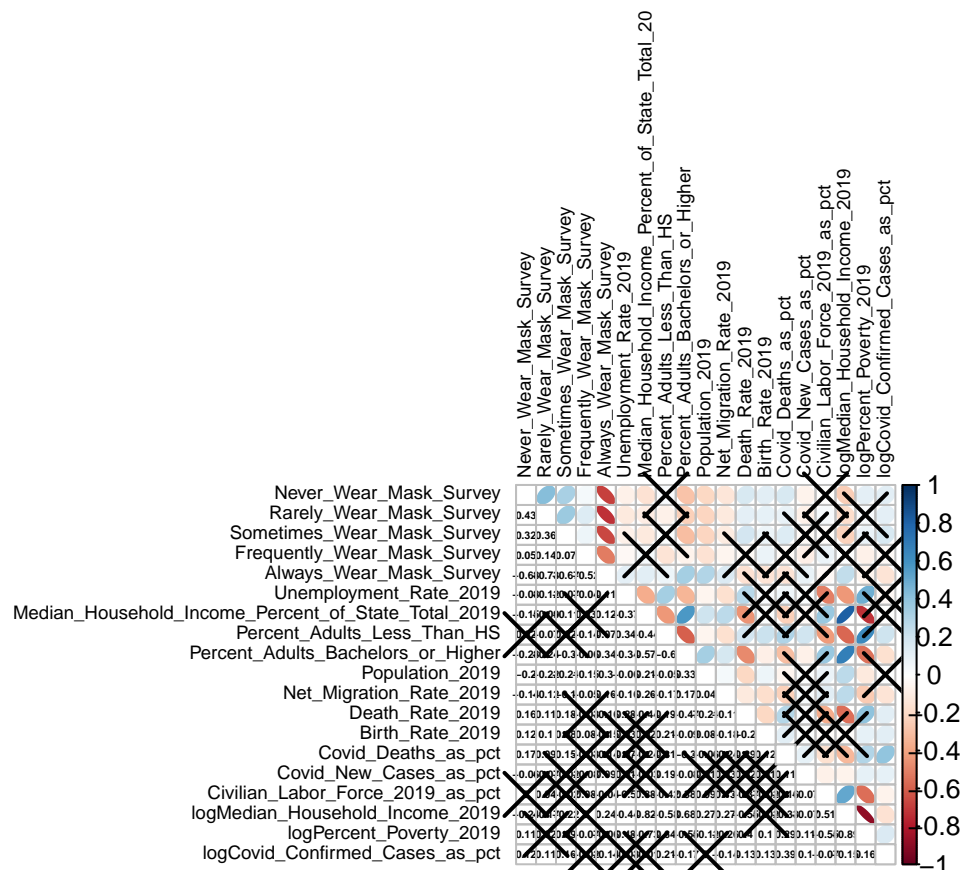
##	Net_Migration_Rate_2019
## Never_Wear_Mask_Survey	-0.13988347
## Rarely_Wear_Mask_Survey	-0.11513583
## Sometimes_Wear_Mask_Survey	-0.10426920
## Frequently_Wear_Mask_Survey	-0.05360013
## Always_Wear_Mask_Survey	0.15800706
## Unemployment_Rate_2019	-0.16283518
## Median_Household_Income_Percent_of_State_Total_2019	0.26481730
## Percent_Adults_Less_Than_HS	-0.16831695
## Percent_Adults_Bachelors_or_Higher	0.17283606
## Population_2019	0.04048487
## Net_Migration_Rate_2019	1.00000000
## Death_Rate_2019	-0.11279934
## Birth_Rate_2019	-0.17606262
## Covid_Deaths_as_pct	-0.23895384
## Covid_New_Cases_as_pct	0.03375891
## Civilian_Labor_Force_2019_as_pct	0.12853296
## logMedian_Household_Income_2019	0.26553925
## logPercent_Poverty_2019	-0.25868300
## logCovid_Confirmed_Cases_as_pct	-0.14091565
##	Death_Rate_2019
## Never_Wear_Mask_Survey	0.16009081
## Rarely_Wear_Mask_Survey	0.11102699
## Sometimes_Wear_Mask_Survey	0.18079613
## Frequently_Wear_Mask_Survey	-0.02818103
## Always_Wear_Mask_Survey	-0.15909794
## Unemployment_Rate_2019	0.27719318
## Median_Household_Income_Percent_of_State_Total_2019	-0.44257122
## Percent_Adults_Less_Than_HS	0.18693289
## Percent_Adults_Bachelors_or_Higher	-0.47306789
## Population_2019	-0.20579228
## Net_Migration_Rate_2019	-0.11279934
## Death_Rate_2019	1.00000000
## Birth_Rate_2019	-0.20201070
## Covid_Deaths_as_pct	0.28656421
## Covid_New_Cases_as_pct	0.02036068
## Civilian_Labor_Force_2019_as_pct	-0.36632867
## logMedian_Household_Income_2019	-0.55212901
## logPercent_Poverty_2019	0.40492204
## logCovid_Confirmed_Cases_as_pct	0.12971573
##	Birth_Rate_2019
## Never_Wear_Mask_Survey	0.12421774
## Rarely_Wear_Mask_Survey	0.10372687
## Sometimes_Wear_Mask_Survey	0.07824170
## Frequently_Wear_Mask_Survey	0.08195148
## Always_Wear_Mask_Survey	-0.14992133
## Unemployment_Rate_2019	0.02863821
## Median_Household_Income_Percent_of_State_Total_2019	0.01921213
## Percent_Adults_Less_Than_HS	0.21065691
## Percent_Adults_Bachelors_or_Higher	-0.08904625
## Population_2019	0.07648350
## Net_Migration_Rate_2019	-0.17606262
## Death_Rate_2019	-0.20201070
## Birth_Rate_2019	1.00000000

## Covid_Deaths_as_pct	0.12287748
## Covid_New_Cases_as_pct	0.01000979
## Civilian_Labor_Force_2019_as_pct	-0.01562853
## logMedian_Household_Income_2019	-0.01761866
## logPercent_Poverty_2019	0.10446736
## logCovid_Confirmed_Cases_as_pct	0.12899745
##	Covid_Deaths_as_pct
## Never_Wear_Mask_Survey	0.16519299
## Rarely_Wear_Mask_Survey	0.09335193
## Sometimes_Wear_Mask_Survey	0.14562233
## Frequently_Wear_Mask_Survey	-0.02761368
## Always_Wear_Mask_Survey	-0.14157596
## Unemployment_Rate_2019	0.06632755
## Median_Household_Income_Percent_of_State_Total_2019	-0.23747279
## Percent_Adults_Less_Than_HS	0.30761178
## Percent_Adults_Bachelors_or_Higher	-0.30057177
## Population_2019	-0.05913505
## Net_Migration_Rate_2019	-0.23895384
## Death_Rate_2019	0.28656421
## Birth_Rate_2019	0.12287748
## Covid_Deaths_as_pct	1.00000000
## Covid_New_Cases_as_pct	0.10830307
## Civilian_Labor_Force_2019_as_pct	-0.16108764
## logMedian_Household_Income_2019	-0.32717512
## logPercent_Poverty_2019	0.29087235
## logCovid_Confirmed_Cases_as_pct	0.39158598
##	Covid_New_Cases_as_pct
## Never_Wear_Mask_Survey	-0.06495969
## Rarely_Wear_Mask_Survey	-0.07158883
## Sometimes_Wear_Mask_Survey	-0.02303206
## Frequently_Wear_Mask_Survey	-0.08147206
## Always_Wear_Mask_Survey	0.09393703
## Unemployment_Rate_2019	0.01246554
## Median_Household_Income_Percent_of_State_Total_2019	-0.02704557
## Percent_Adults_Less_Than_HS	0.19292024
## Percent_Adults_Bachelors_or_Higher	-0.07649416
## Population_2019	0.01081794
## Net_Migration_Rate_2019	0.03375891
## Death_Rate_2019	0.02036068
## Birth_Rate_2019	0.01000979
## Covid_Deaths_as_pct	0.10830307
## Covid_New_Cases_as_pct	1.00000000
## Civilian_Labor_Force_2019_as_pct	-0.07104026
## logMedian_Household_Income_2019	-0.06671465
## logPercent_Poverty_2019	0.10617701
## logCovid_Confirmed_Cases_as_pct	0.10350502
##	Civilian_Labor_Force_2019_as_pct
## Never_Wear_Mask_Survey	-0.001773504
## Rarely_Wear_Mask_Survey	0.035283116
## Sometimes_Wear_Mask_Survey	-0.016213577
## Frequently_Wear_Mask_Survey	0.081487685
## Always_Wear_Mask_Survey	-0.039978566
## Unemployment_Rate_2019	-0.501741424
## Median_Household_Income_Percent_of_State_Total_2019	0.379366543

## Percent_Adults_Less_Than_HS	-0.422758239
## Percent_Adults_Bachelors_or_Higher	0.376973529
## Population_2019	0.086344982
## Net_Migration_Rate_2019	0.128532956
## Death_Rate_2019	-0.366328667
## Birth_Rate_2019	-0.015628530
## Covid_Deaths_as_pct	-0.161087643
## Covid_New_Cases_as_pct	-0.071040262
## Civilian_Labor_Force_2019_as_pct	1.000000000
## logMedian_Household_Income_2019	0.513246563
## logPercent_Poverty_2019	-0.551323967
## logCovid_Confirmed_Cases_as_pct	-0.066296665
##	logMedian_Household_Income_2019
## Never_Wear_Mask_Survey	-0.240717720
## Rarely_Wear_Mask_Survey	-0.173612965
## Sometimes_Wear_Mask_Survey	-0.223060235
## Frequently_Wear_Mask_Survey	0.001936773
## Always_Wear_Mask_Survey	0.240109907
## Unemployment_Rate_2019	-0.438083707
## Median_Household_Income_Percent_of_State_Total_2019	0.817879318
## Percent_Adults_Less_Than_HS	-0.575917423
## Percent_Adults_Bachelors_or_Higher	0.684754417
## Population_2019	0.269169601
## Net_Migration_Rate_2019	0.265539248
## Death_Rate_2019	-0.552129010
## Birth_Rate_2019	-0.017618664
## Covid_Deaths_as_pct	-0.327175124
## Covid_New_Cases_as_pct	-0.066714649
## Civilian_Labor_Force_2019_as_pct	0.513246563
## logMedian_Household_Income_2019	1.000000000
## logPercent_Poverty_2019	-0.885283855
## logCovid_Confirmed_Cases_as_pct	-0.151286701
##	logPercent_Poverty_2019
## Never_Wear_Mask_Survey	0.11330031
## Rarely_Wear_Mask_Survey	0.02027919
## Sometimes_Wear_Mask_Survey	0.08775493
## Frequently_Wear_Mask_Survey	-0.07007921
## Always_Wear_Mask_Survey	-0.05519402
## Unemployment_Rate_2019	0.48249406
## Median_Household_Income_Percent_of_State_Total_2019	-0.72740485
## Percent_Adults_Less_Than_HS	0.63926157
## Percent_Adults_Bachelors_or_Higher	-0.54889770
## Population_2019	-0.11745903
## Net_Migration_Rate_2019	-0.25868300
## Death_Rate_2019	0.40492204
## Birth_Rate_2019	0.10446736
## Covid_Deaths_as_pct	0.29087235
## Covid_New_Cases_as_pct	0.10617701
## Civilian_Labor_Force_2019_as_pct	-0.55132397
## logMedian_Household_Income_2019	-0.88528386
## logPercent_Poverty_2019	1.00000000
## logCovid_Confirmed_Cases_as_pct	0.15835685
##	logCovid_Confirmed_Cases_as_pct
## Never_Wear_Mask_Survey	0.120424610


```
## Rarely_Wear_Mask_Survey 0.110199176
## Sometimes_Wear_Mask_Survey 0.164799288
## Frequently_Wear_Mask_Survey -0.021474934
## Always_Wear_Mask_Survey -0.140271699
## Unemployment_Rate_2019 -0.030763920
## Median_Household_Income_Percent_of_State_Total_2019 -0.005183187
## Percent_Adults_Less_Than_HS 0.207519280
## Percent_Adults_Bachelors_or_Higher -0.165526839
## Population_2019 0.002671028
## Net_Migration_Rate_2019 -0.140915653
## Death_Rate_2019 0.129715734
## Birth_Rate_2019 0.128997450
## Covid_Deaths_as_pct 0.391585980
## Covid_New_Cases_as_pct 0.103505017
## Civilian_Labor_Force_2019_as_pct -0.066296665
## logMedian_Household_Income_2019 -0.151286701
## logPercent_Poverty_2019 0.158356852
## logCovid_Confirmed_Cases_as_pct 1.000000000
```

```
corrplot.mixed(cor(dat), lower.col = "black", upper = "ellipse", tl.col = "black",
               number.cex = .3, tl.pos = "lt", tl.cex = .6,
               p.mat = cor.mtest(dat, conf.level = .95)$p, sig.level = .05)
```



We observe many weak correlations and a few strong correlations.

The masking survey responses are highly correlated. Always masking is inversely correlated with never masking (-0.68), rarely masking (-0.73), sometimes masking (-0.67), and frequently masking (-0.52); this

makes sense considering their linear dependence as limited survey responses. Surprisingly, we do not observe many high correlations with our variables most of interest, rate of COVID-19 cases and COVID-19 death rate.

Education and income are highly correlated, with the percentage of adults with bachelors positively correlated with median household income (0.68) and negatively correlated with adults with less than a high school degree (-0.60). Education and income are also connected to the 2019 death rate, which is inversely correlated with income (-0.55) and adults with higher than a bachelor's degree (-0.47).

Indeed, most of the strong correlations we observe are between variables related to education, employment status, and income. The civilian labor force is positively correlated to those with a bachelor's degree or higher (0.38) and negatively correlated to adults with less than a high school degree (-0.42) and unemployment rate (-0.50). The log of the median household income is positively tied to the income as a percentage of the state total (0.82) and adults with bachelor's degree (0.69) while negatively correlated to employment (-0.44) and adults with less than a high school degree (-0.58). The percentage of the population in poverty is positively correlated to unemployment (0.48) and adults with less than a high school degree (0.64) and negatively correlated to income (-0.88), adults with a bachelor's degree (-0.55), and household income in relation to the state (-0.73).

In short, the strongest observed trends can be summarized with the unsurprising statement: Counties with higher rates of higher education tend to have more people employed, higher median household incomes, and lower death rates. Here, we already begin to see the intersection of wealth and public health.

3) Compute KMO or other measure (i.e. just look at matrix produced above) to comment on suitability of data for factor analysis.

```
paf(as.matrix(dat))$KMO
```

```
## [1] 0.51852
```

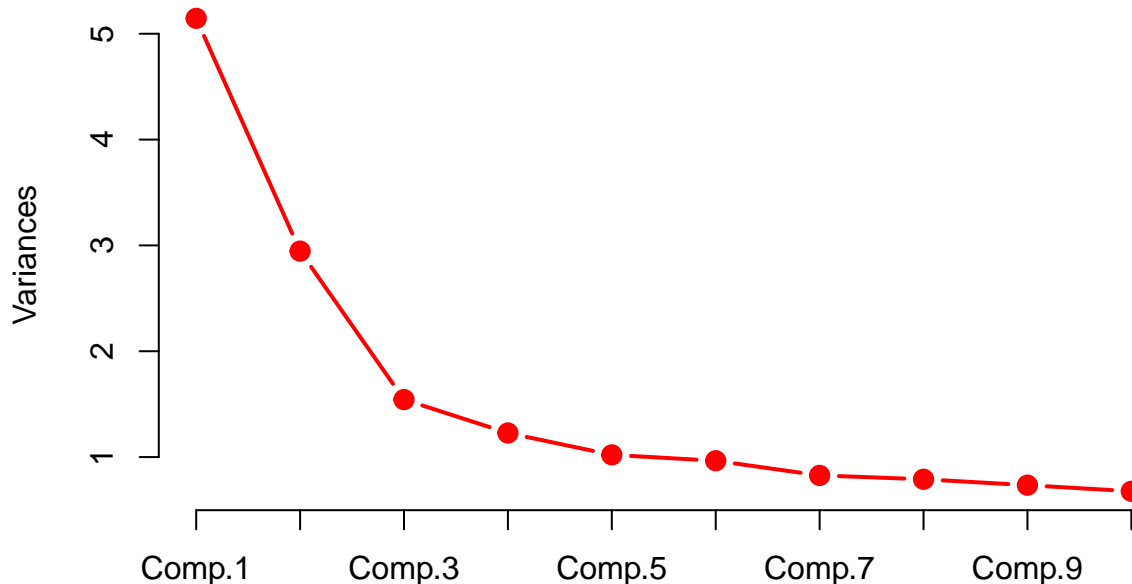
Noting a low KMO of 0.52, we determine that our data is “miserable” for factor analysis. Nevertheless, we continue on! From the matrix above, there are a few very strongly correlated variables with correlations above 0.6 in absolute value - these strong correlations give us hope that factor analysis will produce some interesting results, though it is worth noting that this is probably not the best technique for our data given how complicated county demographics can be.

4) Use Principle Components (or appropriate option in Factor Analysis) to decide on a number of latent factors. You can use Scree Plot, eigenvalue>1, or parallel analysis.

```
pc1 <- princomp(dat, cor=TRUE)

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

Scree Plot of COVID-19 County Data



```
round(pc1$sdev^2, 2)
```

```
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
## 5.15 2.95 1.54 1.23 1.02 0.97 0.83 0.79 0.73 0.68
## Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17 Comp.18 Comp.19
## 0.60 0.55 0.50 0.49 0.47 0.27 0.17 0.06 0.00
```

According to the Eigenvalue > 1 method, the first 5 PC's should be used. According to the scree plot elbow method, the first 3 PC's should be used. Note that parallel analysis, though the preferred method, would not have been appropriate here because the data does not follow a multivariate normal distribution, as has been examined in other psets.

To avoid an unnecessarily complicated model, we will choose to use the scree plot elbow method and assume 3 latent factors.

5) Perform a series of factor analyses using orthogonal models. First, try at least two extraction methods (choose from Principle Components, Principle Axis Factoring, Iterative Principle Components, Maximum Likelihood). Use some method for comparing extraction methods to choose a 'best' method (i.e. RMSR or # residuals greater than .05).

Maximum Likelihood

```
fact1 <- factanal(dat, factors = 3)
fact1
```

```
##
## Call:
## factanal(x = dat, factors = 3)
```

```

##
## Uniquenesses:
##          Never_Wear_Mask_Survey
##                      0.408
##          Rarely_Wear_Mask_Survey
##                      0.397
##          Sometimes_Wear_Mask_Survey
##                      0.457
##          Frequently_Wear_Mask_Survey
##                      0.005
##          Always_Wear_Mask_Survey
##                      0.005
##          Unemployment_Rate_2019
##                      0.724
## Median_Household_Income_Percent_of_State_Total_2019
##                      0.319
##          Percent_Adults_Less_Than_HS
##                      0.588
##          Percent_Adults_Bachelors_or_Higher
##                      0.478
##          Population_2019
##                      0.871
##          Net_Migration_Rate_2019
##                      0.914
##          Death_Rate_2019
##                      0.696
##          Birth_Rate_2019
##                      0.978
##          Covid_Deaths_as_pct
##                      0.876
##          Covid_New_Cases_as_pct
##                      0.980
##          Civilian_Labor_Force_2019_as_pct
##                      0.684
##          logMedian_Household_Income_2019
##                      0.031
##          logPercent_Poverty_2019
##                      0.165
##          logCovid_Confirmed_Cases_as_pct
##                      0.955
##
## Loadings:
##
##          Factor1 Factor2 Factor3
## Never_Wear_Mask_Survey          0.763
## Rarely_Wear_Mask_Survey          0.776
## Sometimes_Wear_Mask_Survey        0.734
## Frequently_Wear_Mask_Survey        0.146    0.985
## Always_Wear_Mask_Survey           -0.917   -0.391
## Unemployment_Rate_2019           -0.503   -0.149
## Median_Household_Income_Percent_of_State_Total_2019  0.818   -0.107
## Percent_Adults_Less_Than_HS       -0.632
## Percent_Adults_Bachelors_or_Higher  0.644   -0.323
## Population_2019                   0.205   -0.267   -0.127
## Net_Migration_Rate_2019           0.252   -0.143

```

```
## Death_Rate_2019 -0.525 0.166
## Birth_Rate_2019 0.135
## Covid_Deaths_as_pct -0.312 0.159
## Covid_New_Cases_as_pct
## Civilian_Labor_Force_2019_as_pct 0.559
## logMedian_Household_Income_2019 0.959 -0.221
## logPercent_Poverty_2019 -0.912
## logCovid_Confirmed_Cases_as_pct -0.127 0.165
##
## Factor1 Factor2 Factor3
## SS loadings 4.320 2.975 1.174
## Proportion Var 0.227 0.157 0.062
## Cumulative Var 0.227 0.384 0.446
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 30739 on 117 degrees of freedom.
## The p-value is 0
```

Principal Axis Factoring

```
# Note that we couldn't get this version of principal axis factoring to converge
# within 2-10 factors
#fact2 <- fa(dat, nfactors = 3, fm = "pa")
```

```
# We also couldn't get iterative PCA to converge
# within 2-10 factors
#fact2 <- fa(dat, nfactors = 3, SMC = FALSE, fm = "pa")
```

```
fact2 <- paf(as.matrix(dat))
summary(fact2)
```

```
## $KMO
## [1] 0.51852
##
## $MSA
## MSA
## Never_Wear_Mask_Survey 0.213193
## Rarely_Wear_Mask_Survey 0.214195
## Sometimes_Wear_Mask_Survey 0.199759
## Frequently_Wear_Mask_Survey 0.085679
## Always_Wear_Mask_Survey 0.347559
## Unemployment_Rate_2019 0.891180
## Median_Household_Income_Percent_of_State_Total_2019 0.884662
## Percent_Adults_Less_Than_HS 0.812795
## Percent_Adults_Bachelors_or_Higher 0.852885
## Population_2019 0.866032
## Net_Migration_Rate_2019 0.873757
## Death_Rate_2019 0.831898
## Birth_Rate_2019 0.632702
## Covid_Deaths_as_pct 0.839948
## Covid_New_Cases_as_pct 0.811235
## Civilian_Labor_Force_2019_as_pct 0.907177
## logMedian_Household_Income_2019 0.805139
## logPercent_Poverty_2019 0.815337
```

```

## logCovid_Confirmed_Cases_as_pct          0.723044
##
## $Bartlett
## [1] 57158
##
## $Communalities
##
## Initial Communalities
## Never_Wear_Mask_Survey          0.999878
## Rarely_Wear_Mask_Survey         0.999864
## Sometimes_Wear_Mask_Survey      0.999876
## Frequently_Wear_Mask_Survey     0.999897
## Always_Wear_Mask_Survey         0.999982
## Unemployment_Rate_2019         0.370446
## Median_Household_Income_Percent_of_State_Total_2019 0.705599
## Percent_Adults_Less_Than_HS     0.628159
## Percent_Adults_Bachelors_or_Higher 0.666718
## Population_2019                0.200889
## Net_Migration_Rate_2019        0.156636
## Death_Rate_2019               0.455598
## Birth_Rate_2019               0.209183
## Covid_Deaths_as_pct            0.285685
## Covid_New_Cases_as_pct         0.062701
## Civilian_Labor_Force_2019_as_pct 0.411031
## logMedian_Household_Income_2019 0.906303
## logPercent_Poverty_2019        0.854351
## logCovid_Confirmed_Cases_as_pct 0.226872
##
## Final Extraction
## Never_Wear_Mask_Survey          0.413916
## Rarely_Wear_Mask_Survey         0.458220
## Sometimes_Wear_Mask_Survey      0.365929
## Frequently_Wear_Mask_Survey     2.303323
## Always_Wear_Mask_Survey         1.180409
## Unemployment_Rate_2019         0.357189
## Median_Household_Income_Percent_of_State_Total_2019 0.634138
## Percent_Adults_Less_Than_HS     0.611274
## Percent_Adults_Bachelors_or_Higher 0.599118
## Population_2019                0.174432
## Net_Migration_Rate_2019        0.127719
## Death_Rate_2019               0.575706
## Birth_Rate_2019               0.347790
## Covid_Deaths_as_pct            0.447267
## Covid_New_Cases_as_pct         0.053236
## Civilian_Labor_Force_2019_as_pct 0.378645
## logMedian_Household_Income_2019 0.928349
## logPercent_Poverty_2019        0.823405
## logCovid_Confirmed_Cases_as_pct 0.383063
##
## $Factor.Loadings
##
##           [,1]      [,2]
## Never_Wear_Mask_Survey -0.324874 -0.39375670
## Rarely_Wear_Mask_Survey -0.256051 -0.48439637
## Sometimes_Wear_Mask_Survey -0.308891 -0.38184804
## Frequently_Wear_Mask_Survey -0.053626 -1.11220755
## Always_Wear_Mask_Survey  0.409274  0.88106856

```

## Unemployment_Rate_2019	-0.440182	0.26888282
## Median_Household_Income_Percent_of_State_Total_2019	0.747840	-0.14933177
## Percent_Adults_Less_Than_HS	-0.618872	0.29715774
## Percent_Adults_Bachelors_or_Higher	0.772346	0.04031601
## Population_2019	0.287662	0.19836866
## Net_Migration_Rate_2019	0.309117	0.04379511
## Death_Rate_2019	-0.572158	0.04859251
## Birth_Rate_2019	-0.104766	-0.10094568
## Covid_Deaths_as_pct	-0.417450	-0.00021177
## Covid_New_Cases_as_pct	-0.078678	0.12273044
## Civilian_Labor_Force_2019_as_pct	0.530760	-0.24974369
## logMedian_Household_Income_2019	0.943735	-0.11550568
## logPercent_Poverty_2019	-0.835418	0.26612020
## logCovid_Confirmed_Cases_as_pct	-0.234423	-0.05431045
##	[,3]	[,4]
## Never_Wear_Mask_Survey	0.3869231	0.0034047
## Rarely_Wear_Mask_Survey	0.3887745	0.0480246
## Sometimes_Wear_Mask_Survey	0.3510762	0.0071961
## Frequently_Wear_Mask_Survey	-1.0280945	-0.0576163
## Always_Wear_Mask_Survey	-0.4651420	-0.0547856
## Unemployment_Rate_2019	-0.2532807	0.1037790
## Median_Household_Income_Percent_of_State_Total_2019	0.1027224	-0.1913860
## Percent_Adults_Less_Than_HS	-0.1427966	-0.3221552
## Percent_Adults_Bachelors_or_Higher	-0.0270326	-0.0113680
## Population_2019	-0.0762332	-0.2088177
## Net_Migration_Rate_2019	-0.0168747	0.1648672
## Death_Rate_2019	0.0098862	0.2853138
## Birth_Rate_2019	0.0442951	-0.4644180
## Covid_Deaths_as_pct	0.0886447	-0.3833833
## Covid_New_Cases_as_pct	-0.0358972	-0.1539614
## Civilian_Labor_Force_2019_as_pct	0.1738052	-0.0565990
## logMedian_Household_Income_2019	0.0784925	-0.1320487
## logPercent_Poverty_2019	-0.1797405	-0.0395066
## logCovid_Confirmed_Cases_as_pct	0.1353748	-0.4537360
##	[,5]	
## Never_Wear_Mask_Survey	-0.060065	
## Rarely_Wear_Mask_Survey	-0.067570	
## Sometimes_Wear_Mask_Survey	0.037428	
## Frequently_Wear_Mask_Survey	0.056072	
## Always_Wear_Mask_Survey	0.131392	
## Unemployment_Rate_2019	-0.127320	
## Median_Household_Income_Percent_of_State_Total_2019	0.073434	
## Percent_Adults_Less_Than_HS	-0.125673	
## Percent_Adults_Bachelors_or_Higher	-0.010704	
## Population_2019	-0.053999	
## Net_Migration_Rate_2019	0.052742	
## Death_Rate_2019	0.405560	
## Birth_Rate_2019	-0.330118	
## Covid_Deaths_as_pct	0.343747	
## Covid_New_Cases_as_pct	0.083609	
## Civilian_Labor_Force_2019_as_pct	0.033995	
## logMedian_Household_Income_2019	0.027810	
## logPercent_Poverty_2019	-0.144202	
## logCovid_Confirmed_Cases_as_pct	0.317737	

```
##
## $RMS
## [1] 0.030702
```

Discussion of Best Method

```
# for maximum likelihood

# get reproduced correlation matrix
repro1 <- fact1$loadings%*%t(fact1$loadings)
# residual correlation matrix
resid1 <- fact1$cor - repro1
# round(resid1, 2)
# get root-mean squared residuals
len <- length(resid1[upper.tri(resid1)])
(RMSR1 <- sqrt(sum(resid1[upper.tri(resid1)]^2)/len))

## [1] 0.066729

# get proportion of residuals greater than 0.05 in absolute value
paste0(round(sum(rep(1, len)[abs(resid1[upper.tri(resid1)])>0.05])/len*100), "%")

## [1] "25%"

# for principal axis factoring
# check summary above as well checking out RMS
summary(fact2)$RMS

## [1] 0.030702
```

For the 3-factor maximum likelihood method, we can see that the root mean squared residual (RMSR) is 0.067 and approximately 25% (i.e., a relatively high percentage) of residuals are greater than 0.05.

For the principal axis factoring (PAF) method, we can see that the RMSR is 0.031.

The lower RMSR of the PAF method suggests that it is the better method. However, note that we ran this PAF method using the `paf` function in the `rela` package. We could not run PAF using the `fa` function in the `psych` package due to a failure to converge. Thus, we remain skeptical about using this more general `fa` function.

6) Once you've chosen an extraction method, try a varimax and/or a quartimax rotation. Pick one of these rotations and discuss the interpretation of the final factors. Make one or more loading plots as appropriate.

We will use factor analysis using maximum likelihood for this analysis.

Let's try a varimax rotation.

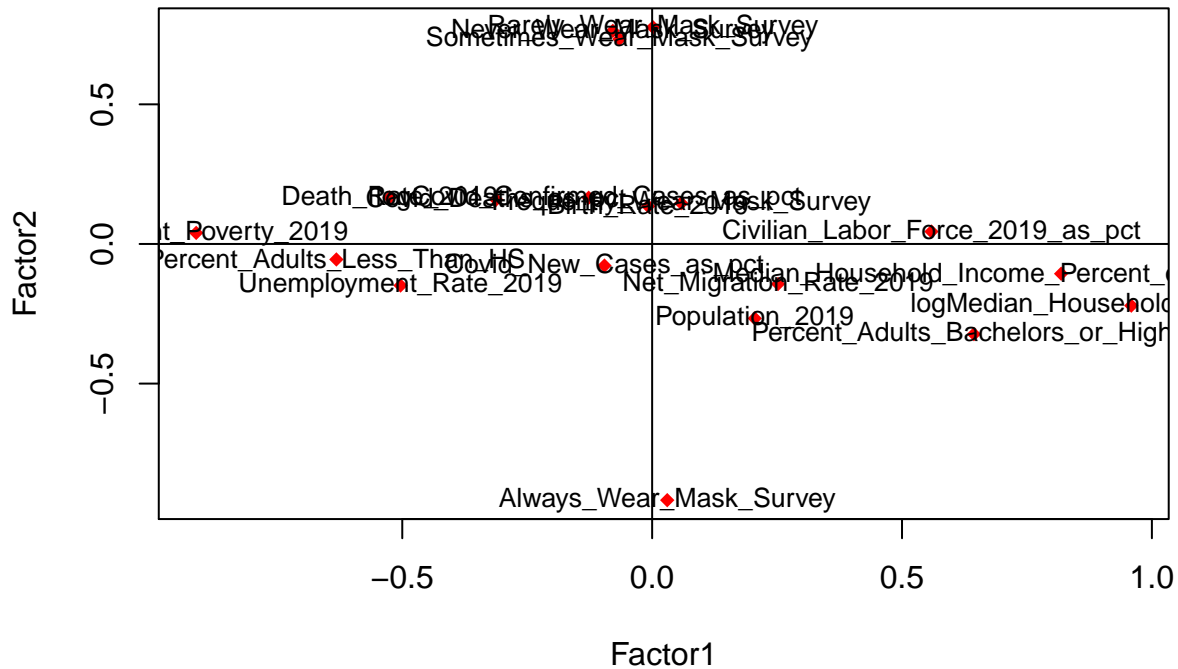
```
fact3 <- factanal(dat, factors = 3, rotation = "varimax")
```

To get the loading plots for this three factor model, we will generate two plots: the first comparing Factor 1 and Factor 2, and the second comparing Factor 1 and Factor 3.

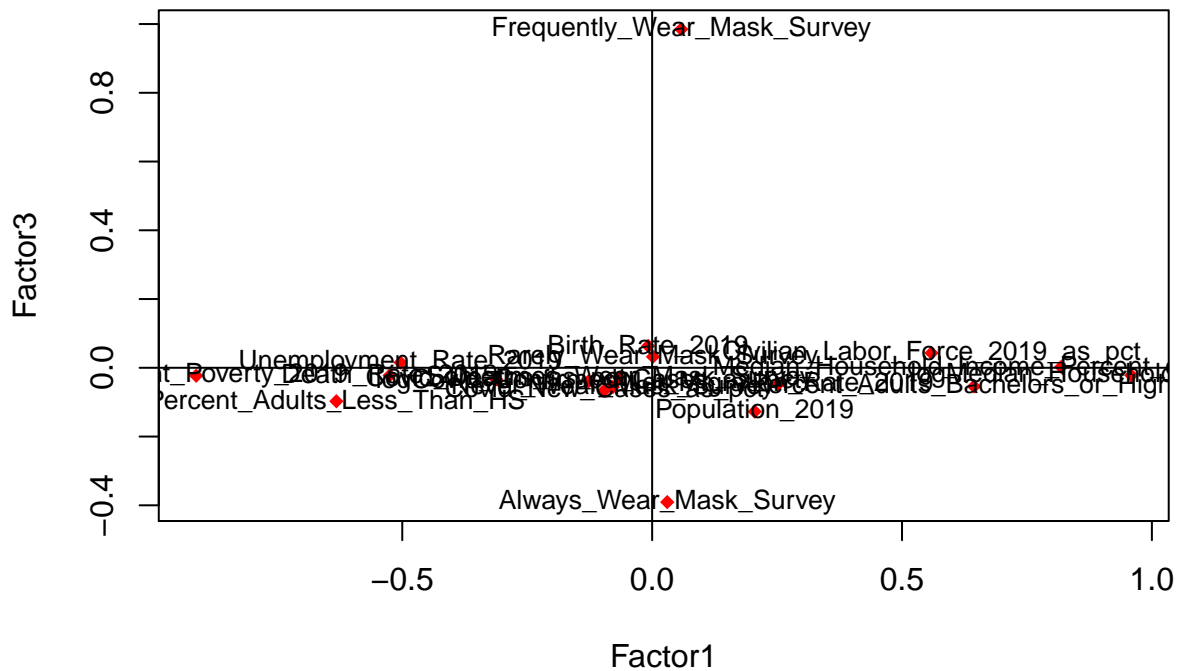
```
# first two factors
plot(fact3$loadings, pch = 18, col = 'red')
abline(h = 0)
```



```
abline(v = 0)
text(fact3$loadings, labels = names(dat), cex = 0.8)
```



```
#factors 1 and 3
plot(fact3$loadings[, c(1,3)], pch = 18, col = 'red')
abline(h = 0)
abline(v = 0)
text(fact3$loadings[, c(1,3)], labels = names(dat), cex = 0.8)
```



Interpreting these plots, we can clearly see that Factor 1 is discriminating between the mask survey responses. The rotation likely aided in allowing the variables `Frequently_Wear_Mask_Survey` and

`Always_Wear_Mask_Survey` to very strictly according to Factor 1 (as opposed to Factor 2 (as seen in the first plot) or Factor 3 (as seen in the second plot)).

From the first plot, we can say that Factor 2 is some sort of economic/development index, with higher values being occupied by variables like `Percent_Adults_Bachelors_or_Higher` and `Civilian_Labor_Force_2019_as_pct`, and with lower values being occupied by variables like `Percent_Adults_Less_Than_HS` and `Unemployment_Rate_2019`. The rotation likely helped reduce variability in the other factor direction.

From the second plot, it seems that Factor 3 is similar to Factor 2 as some sort of economic/development index, although the Factor 3 variables (compared to the Factor 2 variables) demonstrate less variability along Factor 1. The second plot is quite concentrated near the origin, so it is somewhat difficult to interpret. This plot may suggest that Factor 3 is somewhat redundant, considering that it appears to pick up trends in variables similar to those described above for Factor 2. Nevertheless, the scree plot analysis did demonstrate importance to the inclusion of a third factor.

7) Write a short paragraph summarizing your findings.

Using 19 continuous demographic and COVID-19 variables measured on U.S. counties, we determined there to be 3 latent factors (as indicated by a scree plot). Despite this generally reliable determination, two of the three factors proved to be very related in character and potentially redundant. It is unclear why this may have occurred, but it may suggest that the scree plot is not the best way to determine how many factors to use in analyzing our data set. So it goes.

When looking at our results, it is important to consider that we use a few interrelated variables. That's not meant to refer to the moderately-/highly-correlated variables like income and education, but rather those referring to mask-wearing habits. When `Always_Wear_Mask_Survey` increases, the other mask-wearing survey response values must necessarily decrease on net. Perhaps that could contribute to the scree discrepancy.

Or who knows, maybe all our issues trace back to the horrid KMO value.

Ultimately, though, we come out with three or maybe two latent factors in the data, where one is predominantly related to mask habits, and the other is largely a socioeconomic axis. This finding is consistent with what we have seen in previous psets. Principal axis factoring works best for our data, the package limitation described above notwithstanding.