# STAT Project 2: Multivariate Analysis of COVID-19 Daily Cases, Deaths, and Hospitalizations in the City of Chicago

Charles Hwang

5/3/2022

The dataset I chose for this project is the "COVID-19 Daily Cases, Deaths, and Hospitalizations" dataset from the Chicago Data Portal. I chose this dataset because it is updated daily from the City of Chicago's public dashboard and all variables are numeric making it good for quantitative analysis. The dataset includes data for the 785 calendar days from March 1, 2020 through April 24, 2022.[1] There are 58 variables in the dataset: (1) date; (2) total cases, (3) deaths, and (4) hospitalizations; (5-13) cases for ages 0-17, 18-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+, and unknown; (14-16) cases for female, male, and unknown; (17-22) cases for Latinx, Asian, black, white, other, and unknown races[2]; and (23-40) the same groups for deaths and (41-58) hospitalizations.

I read the data into `R` and changed the date variable from character to the `Date` format. I renamed the variables to remove unreadable characters (hyphens) and for convenience and checked the total number of "unknown" values for each category and group (**Figure 1**). Over 96 percent (82,916) of "unknown" values were for race which suggests there may have been trouble or confusion in entering this data. Since the data are time-series, I figured it would be helpful to plot daily (**Figure 2a**) and cumulative (**Figure 3a**) cases, hospitalizations, and deaths over time and the area under the curves for deaths, and since cases are far greater than both hospitalizations and deaths, I plotted the same without them as insets of the original plots (**Figures 2b and 3b**). I also plotted the three variables for each age, gender, and race group[3], which was not too interesting since the graphs all looked the same visually even with different $y$-axes (**Figures 2c through 2e**), but when I plotted the cumulative values with *all groups* on a single graph I was able to more easily compare them. The "0-17" age group had the most cases, generally decreasing with older age groups, but the opposite was true for deaths (**Figure 3c**). The "female" gender group had more cases but "male" had more hospitalizations and deaths (**Figure 3d**). Lastly, "Latinx" had the most cases, but "black" had the most hospitalizations and deaths of any race group by far (**Figure 3e**).

I then computed the correlation and covariance matrices for cases, hospitalizations, and deaths (**Figures 4a and 4b**) and found there are moderately strong positive correlations between the pairings of cases and hospitalizations and hospitalizations and deaths and a weak correlation between cases and deaths. There is a lot of covariance between cases and hospitalizations and considerable covariance between the other two pairs. I also computed correlation and covariance matrices for cases, hospitalizations, and deaths between age, gender, and race for inter-matrix analysis. There is strong to very strong positive correlation for cases between all age groups except unknown and moderate to strong positive correlation for hospitalizations between the

[1] I removed data from April 25 through April 29, 2022, the date I accessed and downloaded the dataset on, because hospitalization data was not yet available. I also removed another row because the date was missing.

[2] For all analyses, I rearranged the columns in the order of cases, hospitalizations, and deaths as that is the order of increasing severity and decreasing quantity. Similarly, I rearranged the race groups in the order of "white", "black", "Latinx", "Asian", "other", and "unknown" in the analysis, as that is the order of decreasing population based on traditional population estimates in the city of Chicago. There should be no change in the results. Although "Latinx" have outpopulated "black" in recent years, I felt it was easier for visual comparison and interpretation to list "black" in the second column.

[3] I excluded the "unknown" groups from race and gender for lack of data and convenience of fitting them in 2 x 4 and 2 x 1 output windows. I tried plotting all groups on a single graph like for the cumulative data, but the groups are very similar and the lines overlapped for each category, making it nearly impossible to reasonably interpret. I also tried plotting base-2 logarithm-transformed cumulative data, but the transformation appeared to be too extreme and unsuitable for the data (since there are multiple "peaks" or "surges").

same (**Figures 4c and 4d**). There are varying levels of positive correlation for deaths (**Figure 4e**), with no correlation when comparing the "0-17" age group to other groups rising progressively to moderately strong positive correlation when doing the same for the "80+" age group. The covariance matrices showed similar trends (**Figures 4f through 4h**), with very high covariance when comparing cases for the "0-17" age group to other groups decreasing progressively to relatively moderate covariance when doing the same for the "80+" age group. Hospitalizations and deaths showed the reverse trend with less covariance involved and nearly no covariance among many pairs of age groups for death. All matrices had little to no covariance for "unknown" age.

Since there were only three levels, I merged cases, hospitalizations, and deaths between gender onto the same matrix and interpreted the three $3 \times 3$ "submatrices" along the diagonal (think of a *Sudoku* grid). I found there was strong to very strong positive correlation between gender groups among all three variables (**Figure 4j**) and nearly perfect positive correlation between female and male cases ($r = 0.99672$). Similar to age, the covariance matrix (**Figure 4k**) showed very high, relatively moderate, and little covariance between gender groups for for the three variables. There were no deaths of either unknown age or gender. Lastly, the matrices between race were similar to age without the progressive change among groups (**Figures 4m through 4s**), except for correlation for deaths, which were moderately positively correlated except for the other and unknown groups which had little to no correlation. I did not see any particularly noteworthy results between any two race groups. Overall, it appeared the results of all the matrices made sense intuitively. I was a bit intrigued at how extreme (both close to 1 and 0) some of the correlation coefficients were, but I figured this may be due to the robust sample size.

I performed principal component analysis (PCA) on total cases, hospitalizations, and deaths and found that the cases variable accounted for over 99.9 percent of the variance (**Figure 5a**). Looking at the signs of the principal components, the first component is a weighted sum, the second component compared cases to hospitalizations and deaths, and the third component compared hospitalizations to cases and deaths. Since there were only three variables, I realized it was only possible for components to be either weighted sums or comparing one variable to the other two. Since the basis of this project is multivariate analysis, I felt it was also meaningful to perform PCA on these variables individually between age and race categories.[4] The **first component for all six analyses** was simply a **weighted sum** of the categories, and all variance proportions progressively decreased in order of input (age increasing or race decreasing by population). For cases and deaths between age groups, there was no practical interpretation as variables were not split contiguously and there was no intuitive comparison. The "0-17" age group accounted for over 95.2 percent of the variance in cases (**Figure 5b**) and nearly 80 percent of the variance in deaths (**Figure 5d**). For hospitalizations between age groups (**Figure 5c**), the third component compared those 0-69 with those 70+ and the fourth was a "60-69" (and unknown) component, but interpretations of other components are complicated for the same reason. The "0-17" age group again accounted for over 86.5 percent of the variance.

For cases between race groups (**Figure 5e**), the second component was a "Latinx" component, the third compared "white", "Asian", and "other" to the inverse, the fourth component compared "white", "black", and "Latinx" (the three most populous race groups) to the inverse, the fifth was a "Latinx"/"other" component, and the sixth compared "black", "Asian", and "other" (the three least populous race groups in the dataset) to the inverse. "White" accounted for over 91.6 percent of the variance. For hospitalizations between race groups (**Figure 5f**), the second component was a "black" component, the third was a "black"/"Latinx" component, and the fourth through sixth components compared the $\binom{3}{2} = 3$ possible pairs of "Asian", "other", and "unknown" to their inverses. "White" again accounted for nearly 90 percent of the variance. Lastly, for deaths between race groups (**Figure 5g**), the second component was a "black" component, the third compared "black", "Latinx", and "Asian" (the three most populous non-white race groups) to the inverse, the fourth was a "Asian"/"unknown" component, the fifth compared "white", "black", and "Latinx" to the inverse, and the sixth compared "white", "Asian", and "other" to the inverse. "White" accounted for over 80.2 percent of the variance. Looking at the cumulative plots (**Figures 3c and 3e**), these components made sense intuitively and accurately reflect the importance and proportions of each group.

I could not perform multivariate analysis of variance (MANOVA) because the data are time-series which violates

[4] I excluded gender because there are only three levels and the same issue as before would arise. I also did not feel it was practical to interpret components if the three variables were combined.

the randomness and independence assumptions, and additionally several variables exhibit multicollinearity and do not appear to be multivariate normal. Thus, I performed factor analysis on the same six subsets, again excluding gender because I did not feel it was practical to interpret the factors of combined variables. For each subset, I chose the largest possible $m$ that satisfies $(p-m)^2 - p - m \geq 0$ (where $p$ is the number of variables in the subset) then performed an additional analysis with a smaller value of $m$ after reviewing the initial factors (and the $p$-value from the accompanying $\chi^2$ goodness-of-fit test, if applicable). Of course, the selection of $m$ is subjective, but I did not feel that testing every value of $m$ would be useful so I just chose the maximum and "best" values. I also verified that $\Sigma = \mathbf{LL}' + \Psi$ (Example 9.1, pages 484-485) and wrote the code to produce factor scores from the weighted least squares and regression methods for each analysis, but did not print the results because I felt it would be impractical to include in the Appendix and not very meaningful to the overall analysis.

For both analyses of cases and hospitalizations between age groups (**Figures 6a through 6d**), the first factor was for old age and the second was for young age (as the loadings were in increasing and decreasing orders respectively), and the analyses for cases additionally had third and fourth factors for those 0-17. For deaths between age groups, the first factor was for old age, but the others were difficult to interpret as they were not in the order of age groups and there was no intuitive comparison (**Figures 6e and 6f**). For both analyses of cases between race groups (**Figure 6g and 6h**), the first factor was for "white", "black", "other", and "unknown" and the second factor was for "Latinx", and the $m = 3$ analysis additionally had an "Asian" factor. For hospitalizations between race groups ($m = 3$), the first factor was for "white" and "black", the second was for "unknown", and the third was for "Latinx" (**Figure 6j**). The $m = 2$ analysis had a general factor (every group except "unknown") and a "white", "black", and "unknown" factor (**Figure 6k**). Lastly, for both analyses of deaths between race groups (**Figure 6m and 6n**), the first factor was for "Latinx", "black", and "white", and the $m = 3$ analysis additionally had "Asian" and "white" factors. These factors also make sense intuitively based on the previous results of the plots and PCA.

Additional multivariate analyses like clustering, canonical correlation analysis (CCA), etc. could be performed if I had more time with this project. I could also perform time-series analysis incorporating multivariate topics like multiple regression models (Section 7.10, page 413) if I had more experience in these subjects (I believe this is the first university project for which I have used time-series data). Although the gender groups appeared to be very highly correlated, there could be additional analysis using this type of model in place of PCA and factor analysis. It does not appear to be possible with this dataset, but I would also be interested in investigations of the interactions between age, gender, and race and compare the three outcome variables for various sets of groups. If truly motivated, I could treat the unknown values (**Figure 1**) as missing data and use multiple imputation by chained equations (MICE), machine learning, or some similar method to predict the groups they belonged to. Either way, I believe I was able to rectify the mistakes of dataset selection from the previous project by vetting and choosing a robust dataset which made it easier to manipulate and interpret. I had some initial doubts I would be able to successfully analyze the dataset because it was so recent and may be more complex, but the variables were subsetted by age, gender, and race which was quite helpful. There did not appear to be any problems with performing any of the analyses.

Overall, I felt this was a good application of multivariate statistical methods. I found it interesting that the earliest date in the dataset was March 1, 2020, as the earliest recorded deaths in the United States were in January 2020 and the earliest recorded cases were in 2019. After some brief research, it appears March 1, 2020 was when the City of Chicago began recording these data or is at least the extent to which they have made them publicly available. I was able to choose a relevant and interesting dataset and perform thorough multivariate analyses using the topics learned in the second half of the course.

## Appendix

### Figure 1

```r
rm(list=ls())
C<-read.csv("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/COVID-19
C$Date<-as.Date(C$Date,"%m/%d/%Y")
```

```r
C<-na.omit(C[order(C$Date),])
row.names(C)<-1:nrow(C)
names(C)<-c("Date","C","D","H","C17","18C29","30C39","40C49","50C59","60C69","70C79","80C","C?A","CF","(
sapply(C[,c("C?A","H?A","D?A","C?G","H?G","D?G","C?R","H?R","D?R")],sum) # "Unknown"
```

```
##    C?A    H?A    D?A    C?G    H?G    D?G    C?R    H?R    D?R
##    107      7      0   2469      6      0  82222    680     14
```

**Figure 2a**

```r
plot(C$Date,C$C,type="l",xlab="",ylab="",las=2,xaxt="n",main="COVID-19 Cases, Hospitalizations, and Dea
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$H,type="l",col="blue")
points(C$Date,C$D,type="l",col="red")
polygon(C$Date,C$D,col="orange",lty=0)
legend("topleft",c("Confirmed Cases","Hospitalizations","Confirmed Deaths"),col=c("black","blue","red")
```

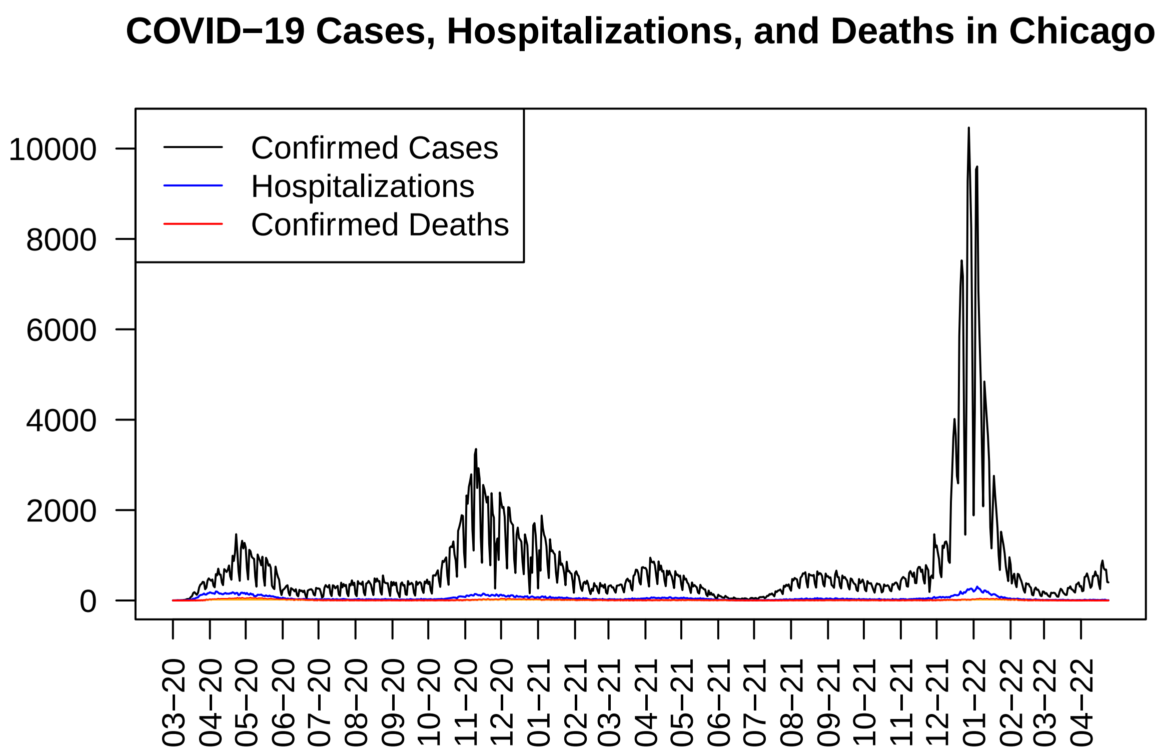



**Figure 2b**

```r
plot(C$Date,C$H,type="l",xlab="",ylab="",las=2,col="blue",xaxt="n",main="COVID-19 Hospitalizations and D
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$D,type="l",col="red")
segments(as.numeric(C$Date)[1],0,as.numeric(C$Date)[nrow(C)])
polygon(C$Date,C$D,col="orange",lty=0)
legend("topleft",c("Hospitalizations","Confirmed Deaths"),col=c("blue","red"),lty=1)
```

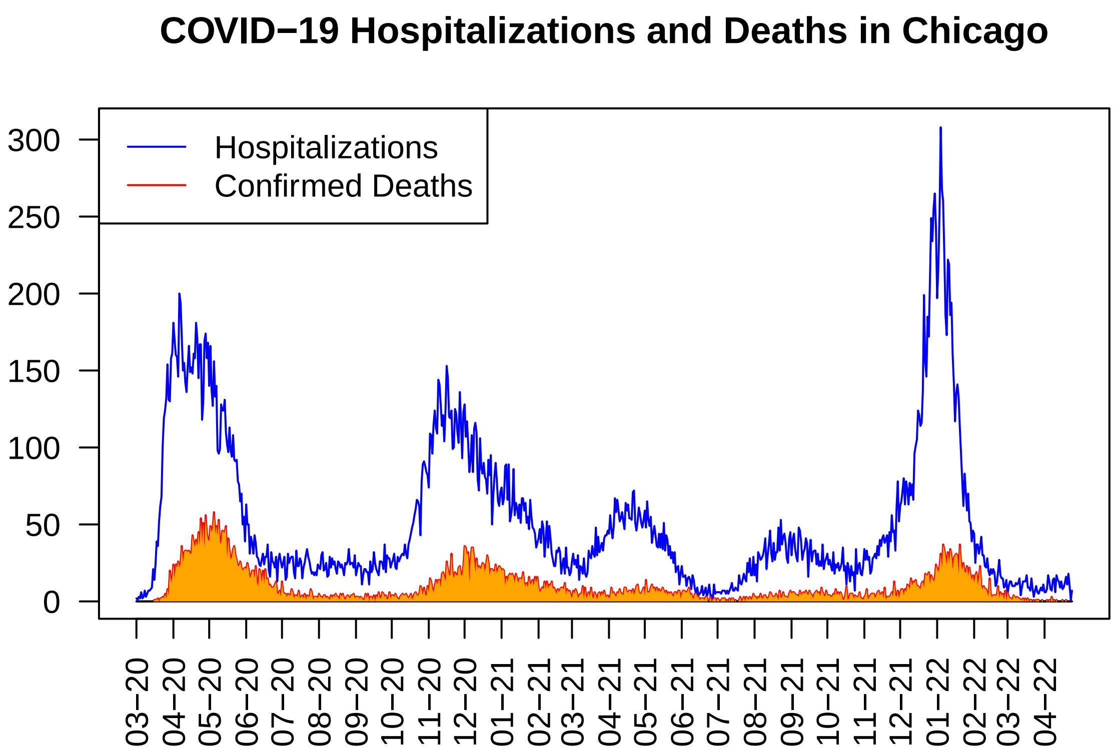



**Figure 2c**

```
par(mfrow=c(2,4))
plot(C$Date,C$C17,type="l",xlab="",ylab="",las=2,xaxt="n",main="Ages 0-17")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$H17,type="l",col="blue")
points(C$Date,C$D17,type="l",col="red")
plot(C$Date,C$`18C29`,type="l",xlab="",ylab="",las=2,xaxt="n",main="Ages 18-29")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$`18H29`,type="l",col="blue")
points(C$Date,C$`18D29`,type="l",col="red")
plot(C$Date,C$`30C39`,type="l",xlab="",ylab="",las=2,xaxt="n",main="Ages 30-39")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$`30H39`,type="l",col="blue")
points(C$Date,C$`30D39`,type="l",col="red")
plot(C$Date,C$`40C49`,type="l",xlab="",ylab="",las=2,xaxt="n",main="Ages 40-49")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$`40H49`,type="l",col="blue")
points(C$Date,C$`40D49`,type="l",col="red")
plot(C$Date,C$`50C59`,type="l",xlab="",ylab="",las=2,xaxt="n",main="Ages 50-59")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$`50H59`,type="l",col="blue")
points(C$Date,C$`50D59`,type="l",col="red")
plot(C$Date,C$`60C69`,type="l",xlab="",ylab="",las=2,xaxt="n",main="Ages 60-69")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$`60H69`,type="l",col="blue")
points(C$Date,C$`60D69`,type="l",col="red")
plot(C$Date,C$`70C79`,type="l",xlab="",ylab="",las=2,xaxt="n",main="Ages 70-79")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
```

```
points(C$Date,C$`70H79`,type="l",col="blue")
points(C$Date,C$`70D79`,type="l",col="red")
plot(C$Date,C$`80C`,type="l",xlab="",ylab="",las=2,xaxt="n",main="Ages 80+")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$`80H`,type="l",col="blue")
points(C$Date,C$`80D`,type="l",col="red")
```

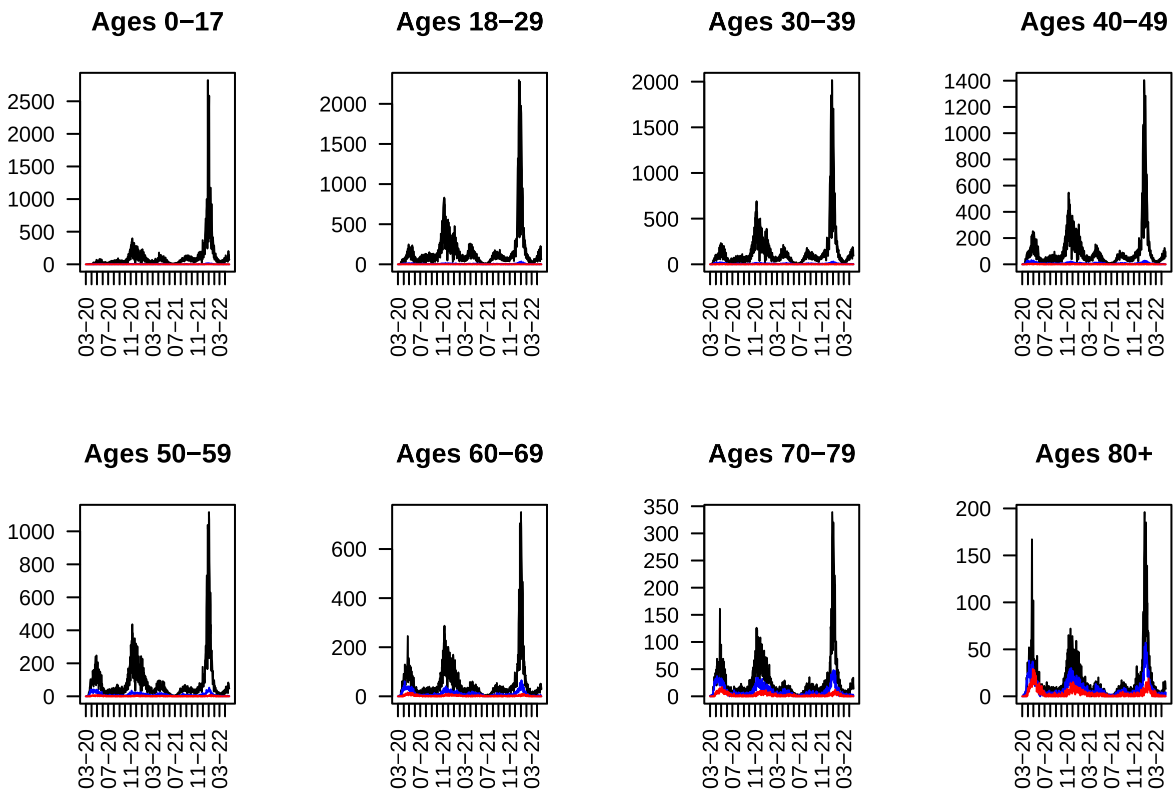
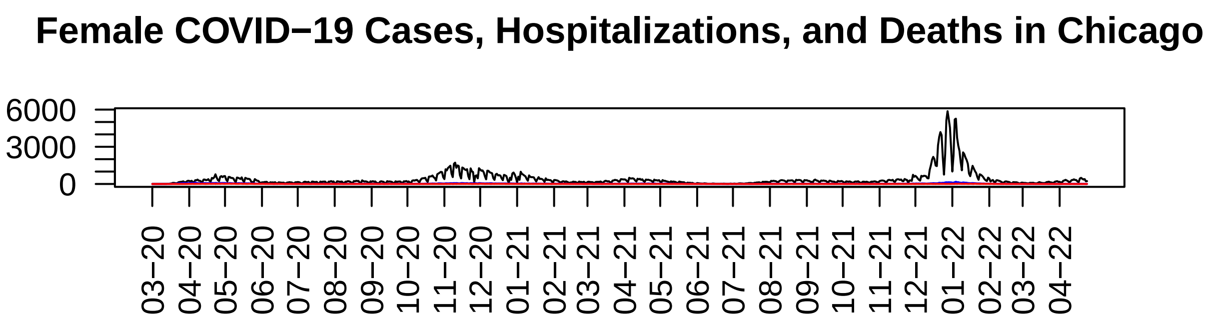



**Figure 2d**

```
par(mfrow=c(2,1))
plot(C$Date,C$CF,type="l",xlab="",ylab="",las=2,xaxt="n",main="Female COVID-19 Cases, Hospitalizations,
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$HF,type="l",col="blue")
points(C$Date,C$DF,type="l",col="red")
plot(C$Date,C$CM,type="l",xlab="",ylab="",las=2,xaxt="n",main="Male COVID-19 Cases, Hospitalizations, an
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$HM,type="l",col="blue")
points(C$Date,C$DM,type="l",col="red")
```



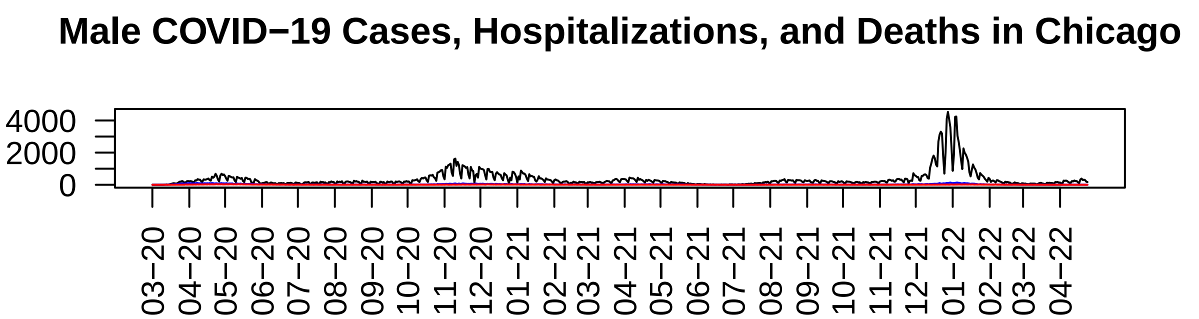




**Figure 2e**

```
par(mfrow=c(2,3))
plot(C$Date,C$CW,type="l",xlab="",ylab="",las=2,xaxt="n",main="White")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$HW,type="l",col="blue")
points(C$Date,C$DW,type="l",col="red")
legend("topleft",c("Cases","Hospitalizations","Deaths"),cex=0.85,col=c("black","blue","red"),lty=1)
plot(C$Date,C$CB,type="l",xlab="",ylab="",las=2,xaxt="n",main="Black")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$HB,type="l",col="blue")
points(C$Date,C$DB,type="l",col="red")
plot(C$Date,C$CL,type="l",xlab="",ylab="",las=2,xaxt="n",main="Latinx")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$HL,type="l",col="blue")
points(C$Date,C$DL,type="l",col="red")
plot(C$Date,C$CA,type="l",xlab="",ylab="",las=2,xaxt="n",main="Asian")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$HA,type="l",col="blue")
points(C$Date,C$DA,type="l",col="red")
plot(C$Date,C$Co,type="l",xlab="",ylab="",las=2,xaxt="n",main="Other")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$Ho,type="l",col="blue")
points(C$Date,C$Do,type="l",col="red")
plot(C$Date,C$`C?R`,type="l",xlab="",ylab="",las=2,xaxt="n",main="Unknown")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,C$`H?R`,type="l",col="blue")
points(C$Date,C$`D?R`,type="l",col="red")
```

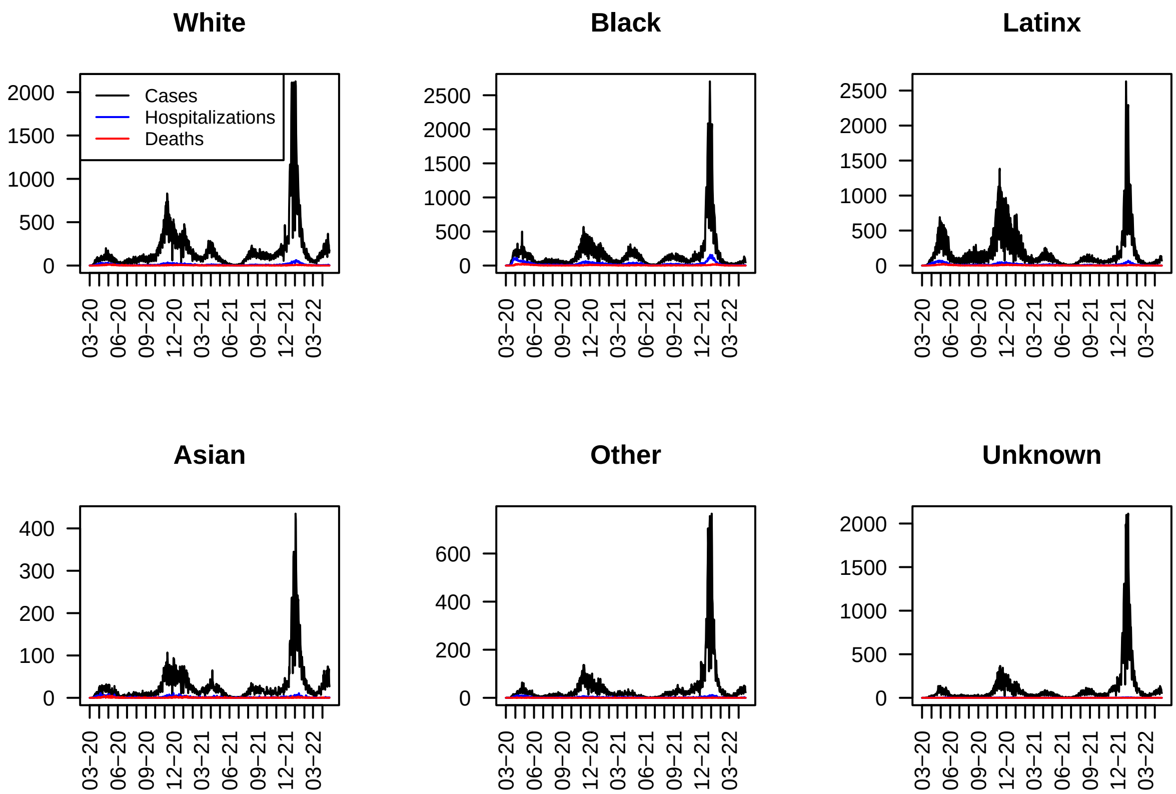



**Figure 3a**

```
plot(C$Date,cumsum(C$C),type="l",xlab="",ylab="",xaxt="n",yaxt="n",main="Cumulative Chicago COVID-19 Cas
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
axis(2,axTicks(2),format(axTicks(2),scientific=FALSE),las=2)
points(C$Date,cumsum(C$H),type="l",col="blue")
points(C$Date,cumsum(C$D),type="l",col="red")
polygon(c(C$Date,max(C$Date)),c(cumsum(C$D),0),col="orange",lty=0)
legend("topleft",c("Cumulative Confirmed Cases","Cumulative Hospitalizations","Cumulative Confirmed Deat
```

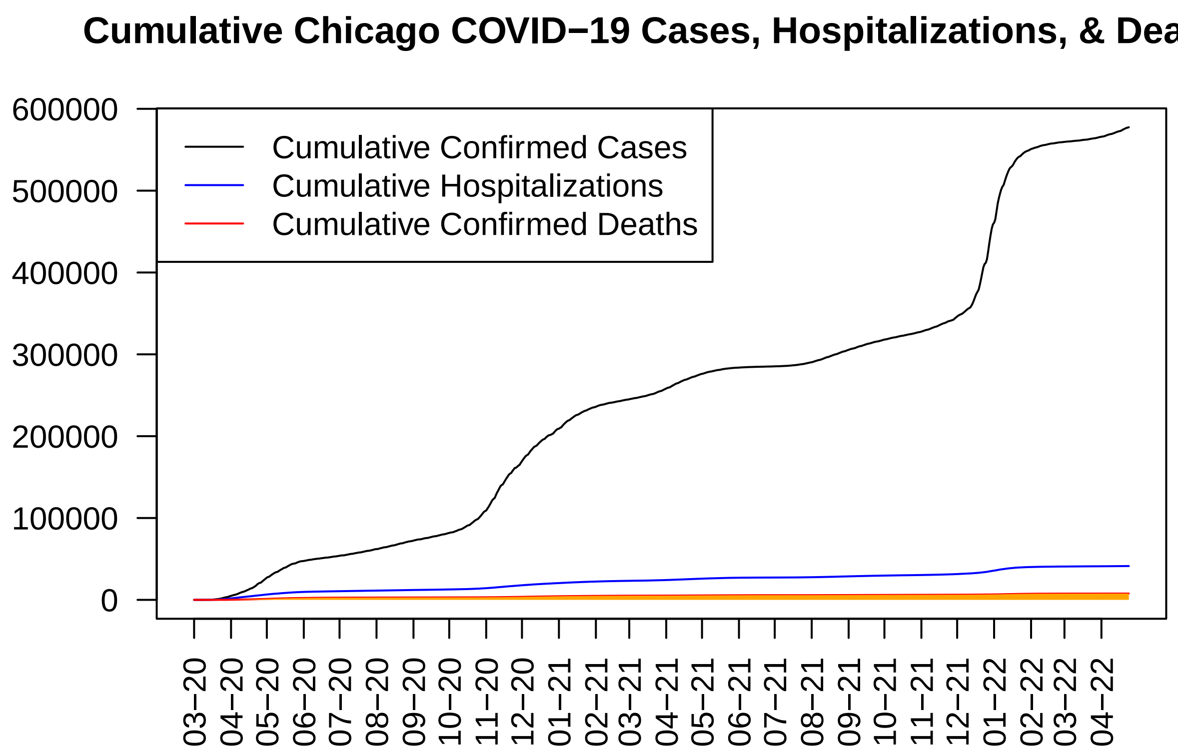



**Figure 3b**

```
plot(C$Date,cumsum(C$H),type="l",xlab="",ylab="",las=2,col="blue",xaxt="n",main="Cumulative Chicago COVI
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,cumsum(C$D),type="l",col="red")
polygon(c(C$Date,max(C$Date)),c(cumsum(C$D),0),col="orange",lty=0)
legend("topleft",c("Cumulative Hospitalizations","Cumulative Confirmed Deaths"),col=c("blue","red"),lty=
```

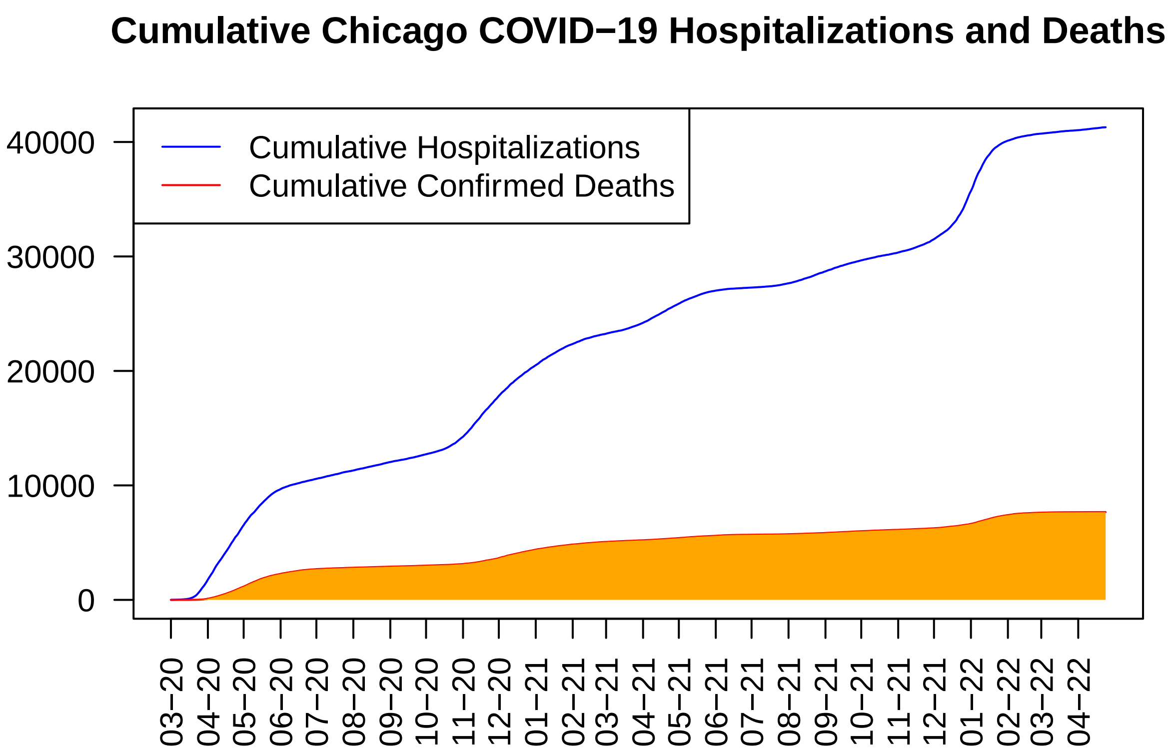



**Figure 3c**

```
par(mfrow=c(1,3))
plot(C$Date,cumsum(C$`18C29`),type="l",xlab="",ylab="",las=2,xaxt="n",main="Cumulative Cases")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,cumsum(C$C17),type="l",col="skyblue")
points(C$Date,cumsum(C$`30C39`),type="l",col="blue")
points(C$Date,cumsum(C$`40C49`),type="l",col="green4")
points(C$Date,cumsum(C$`50C59`),type="l",col="green")
points(C$Date,cumsum(C$`60C69`),type="l",col="yellow")
points(C$Date,cumsum(C$`70C79`),type="l",col="orange")
points(C$Date,cumsum(C$`80C`),type="l",col="red")
points(C$Date,cumsum(C$`C?A`),type="l",col="gray")
legend("topleft",c("Ages 0-17","Ages 18-29","Ages 30-39","Ages 40-49","Ages 50-59","Ages 60-69","Ages 7
plot(C$Date,cumsum(C$`60H69`),col="yellow",type="l",xlab="",ylab="",las=2,xaxt="n",main="Cumulative Hos
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,cumsum(C$H17),type="l",col="skyblue")
points(C$Date,cumsum(C$`18H29`),type="l")
points(C$Date,cumsum(C$`30H39`),type="l",col="blue")
points(C$Date,cumsum(C$`40H49`),type="l",col="green4")
points(C$Date,cumsum(C$`50H59`),type="l",col="green")
points(C$Date,cumsum(C$`70H79`),type="l",col="orange")
points(C$Date,cumsum(C$`80H`),type="l",col="red")
points(C$Date,cumsum(C$`H?A`),type="l",col="gray")
legend("topleft",c("0-17","18-29","30-39","40-49","50-59","60-69","70-79","80+","Unk."),col=c("black","
plot(C$Date,cumsum(C$`80D`),col="red",type="l",xlab="",ylab="",las=2,xaxt="n",main="Cumulative Deaths")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,cumsum(C$D17),type="l",col="skyblue")
points(C$Date,cumsum(C$`18D29`),type="l")
```

```
points(C$Date,cumsum(C$`30D39`),type="l",col="blue")
points(C$Date,cumsum(C$`40D49`),type="l",col="green4")
points(C$Date,cumsum(C$`50D59`),type="l",col="green")
points(C$Date,cumsum(C$`60D69`),type="l",col="yellow")
points(C$Date,cumsum(C$`70D79`),type="l",col="orange")
points(C$Date,cumsum(C$`D?A`),type="l",col="gray")
legend("topleft",c("0-17","18-29","30-39","40-49","50-59","60-69","70-79","80+","Unk."),col=c("black","s
```

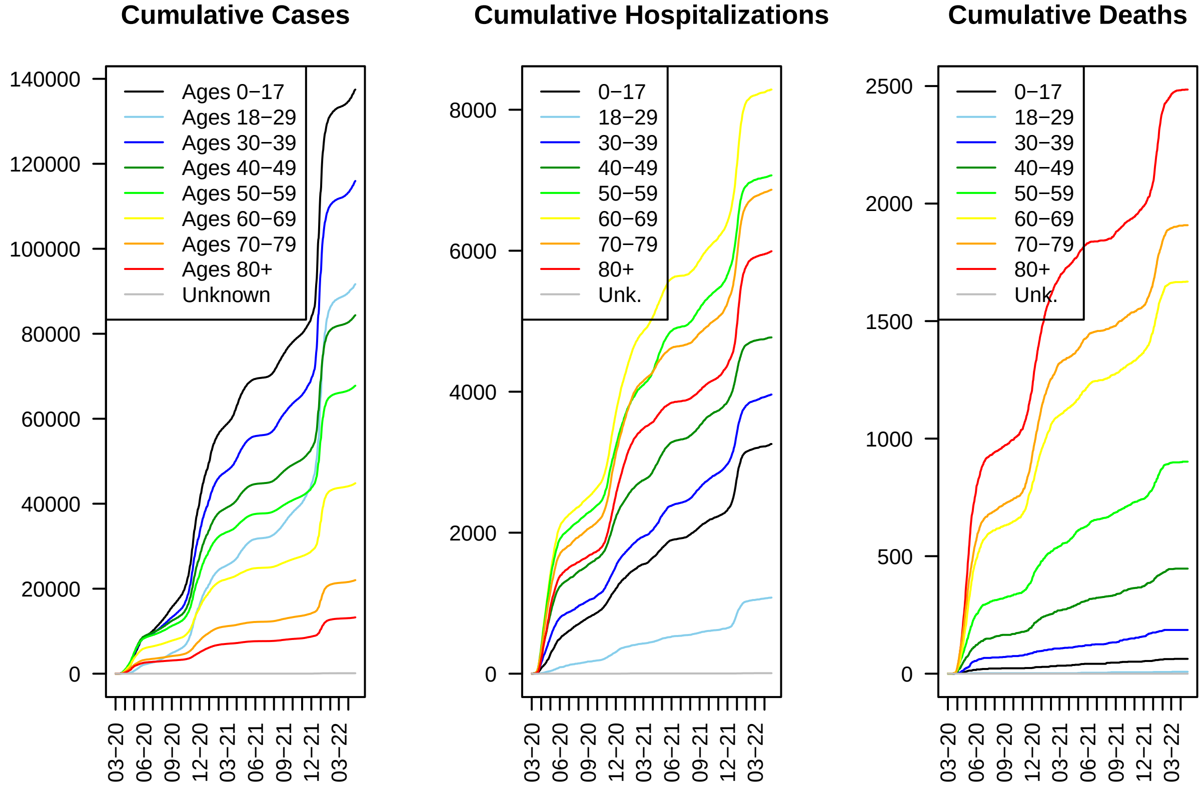



**Figure 3d**

```
par(mfrow=c(1,2))
plot(C$Date,cumsum(C$CF),type="l",xlab="",ylab="",las=2,xaxt="n",main="Cumulative Cases, Gender")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,cumsum(C$CM),type="l",col="red")
points(C$Date,cumsum(C$`C?G`),type="l",col="gray")
legend("topleft",c("Female","Male","Unknown"),col=c("black","red","gray"),lty=1)
plot(C$Date,cumsum(C$HM),col="red",type="l",xlab="",ylab="",las=2,xaxt="n",main="Cum. Hosp./Deaths, Gend
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,cumsum(C$HF),type="l")
points(C$Date,cumsum(C$`H?G`),type="l",col="gray")
points(C$Date,cumsum(C$DM),type="l",col="orange")
points(C$Date,cumsum(C$DF),type="l",col="blue")
points(C$Date,cumsum(C$`D?G`),type="l",col="gray")
legend("topleft",c("H (M)","H (F)","D (M)","D (F)","Unk."),col=c("red","black","orange","blue","gray"),l
```

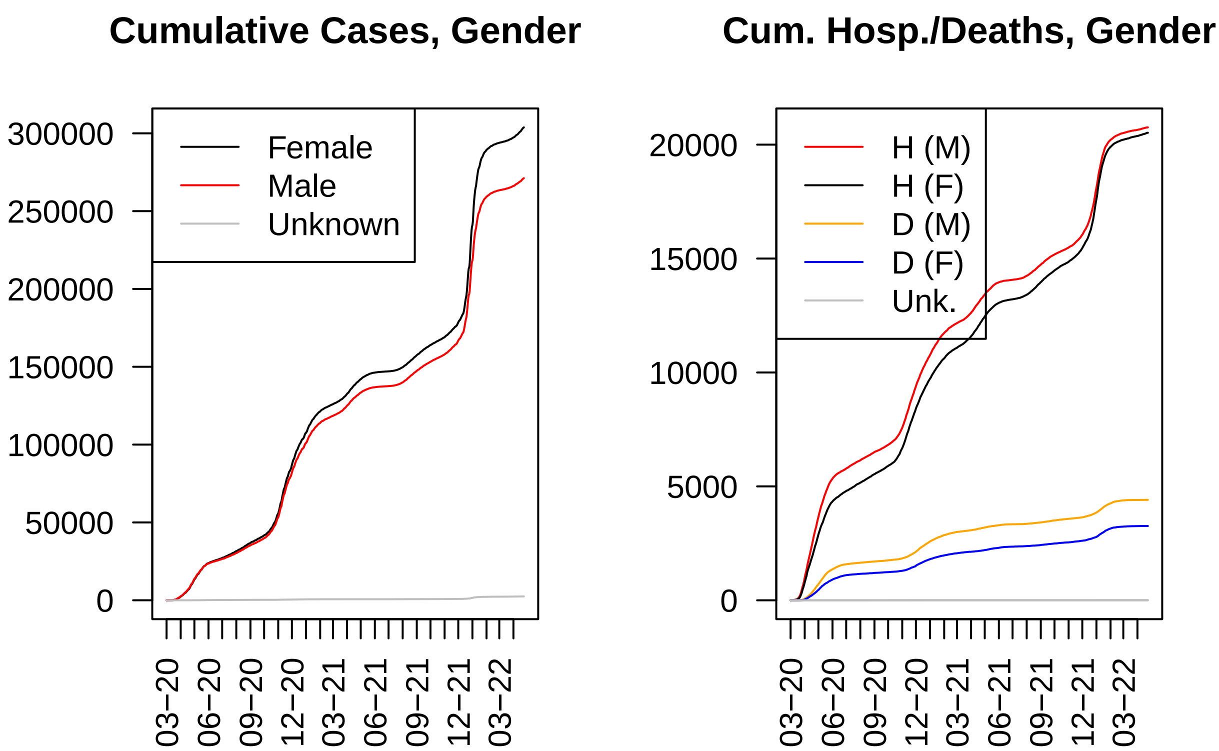



**Figure 3e**

```
par(mfrow=c(1,3))
plot(C$Date,cumsum(C$CL),type="l",xlab="",ylab="",las=2,xaxt="n",main="Cumulative Cases")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,cumsum(C$CW),type="l",col="tan")
points(C$Date,cumsum(C$CB),type="l",col="brown")
points(C$Date,cumsum(C$CA),type="l",col="yellow")
points(C$Date,cumsum(C$Co),type="l",col="blue")
points(C$Date,cumsum(C$`C?R`),type="l",col="red")
legend("topleft",c("White","Black","Latinx","Asian","Other","Unknown"),col=c("tan","brown","black","yel
plot(C$Date,cumsum(C$HB),col="brown",type="l",xlab="",ylab="",las=2,xaxt="n",main="Cumulative Hospitali
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,cumsum(C$HW),type="l",col="tan")
points(C$Date,cumsum(C$HL),type="l")
points(C$Date,cumsum(C$HA),type="l",col="yellow")
points(C$Date,cumsum(C$Ho),type="l",col="blue")
points(C$Date,cumsum(C$`H?R`),type="l",col="red")
legend("topleft",c("White","Black","Latinx","Asian","Other","Unknown"),col=c("tan","brown","black","yel
plot(C$Date,cumsum(C$DB),col="brown",type="l",xlab="",ylab="",las=2,xaxt="n",main="Cumulative Deaths")
axis.Date(1,at=seq(min(C$Date),max(C$Date),by="1 mon"),format="%m-%y",las=2)
points(C$Date,cumsum(C$DW),type="l",col="tan")
points(C$Date,cumsum(C$DL),type="l")
points(C$Date,cumsum(C$DA),type="l",col="yellow")
points(C$Date,cumsum(C$Do),type="l",col="blue")
points(C$Date,cumsum(C$`D?R`),type="l",col="red")
legend("topleft",c("White","Black","Latinx","Asian","Other","Unknown"),col=c("tan","brown","black","yel
```

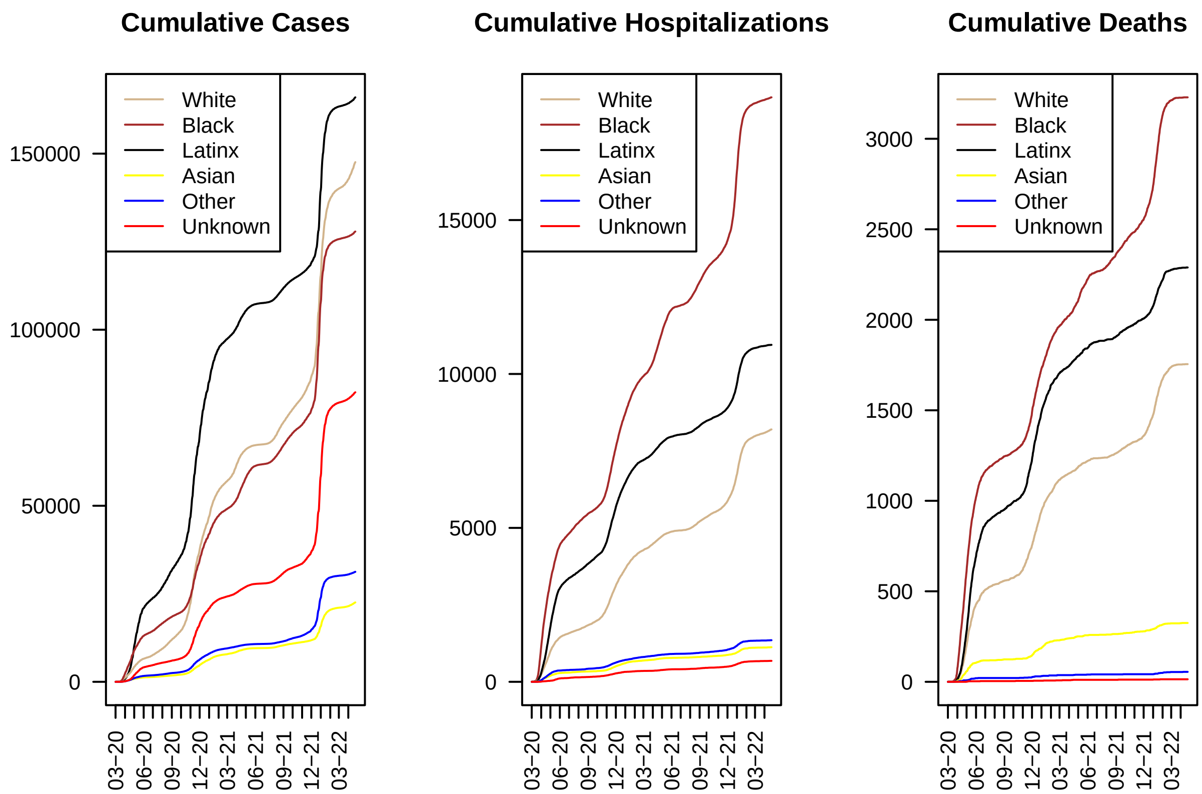



**Figure 4a**

```
cor(C[,c("C","H","D")])
```

```
##           C         H         D
## C 1.0000000 0.7384315 0.3432732
## H 0.7384315 1.0000000 0.7299803
## D 0.3432732 0.7299803 1.0000000
```

**Figure 4b**

```
cov(C[,c("C","H","D")])
```

```
##            C          H         D
## C 1378835.39 43811.9074 4349.0405
## H   43811.91  2553.0041  397.9552
## D    4349.04   397.9552  116.4111
```

**Figure 4c**

```
CA<-C[,c("C17","18C29","30C39","40C49","50C59","60C69","70C79","80C","C?A")]
round(cor(CA),5)
```

```
##           C17   18C29   30C39   40C49   50C59   60C69   70C79     80C     C?A
## C17   1.00000 0.89105 0.90527 0.90668 0.90493 0.89823 0.87457 0.80317 0.51281
## 18C29 0.89105 1.00000 0.99457 0.97528 0.94391 0.91342 0.86396 0.78553 0.39643
## 30C39 0.90527 0.99457 1.00000 0.98327 0.95553 0.92859 0.88342 0.80878 0.41629
## 40C49 0.90668 0.97528 0.98327 1.00000 0.98658 0.96903 0.93539 0.86891 0.41907
## 50C59 0.90493 0.94391 0.95553 0.98658 1.00000 0.98854 0.96229 0.90886 0.42641
```

```
## 60C69 0.89823 0.91342 0.92859 0.96903 0.98854 1.00000 0.97953 0.93561 0.42979
## 70C79 0.87457 0.86396 0.88342 0.93539 0.96229 0.97953 1.00000 0.95677 0.40012
## 80C   0.80317 0.78553 0.80878 0.86891 0.90886 0.93561 0.95677 1.00000 0.38885
## C?A   0.51281 0.39643 0.41629 0.41907 0.42641 0.42979 0.40012 0.38885 1.00000
```

**Figure 4d**

```
HA<-C[,c("H17","18H29","30H39","40H49","50H59","60H69","70H79","80H","H?A")]
round(cor(HA),5)
```

```
##           H17   18H29   30H39   40H49   50H59   60H69   70H79     80H     H?A
## H17   1.00000 0.72336 0.62718 0.49889 0.50706 0.56965 0.55230 0.60440 0.10335
## 18H29 0.72336 1.00000 0.72812 0.63462 0.65465 0.69680 0.69040 0.73100 0.09188
## 30H39 0.62718 0.72812 1.00000 0.73288 0.77387 0.80770 0.76756 0.75979 0.07309
## 40H49 0.49889 0.63462 0.73288 1.00000 0.83605 0.83213 0.81174 0.78068 0.03805
## 50H59 0.50706 0.65465 0.77387 0.83605 1.00000 0.88401 0.87045 0.83564 0.04966
## 60H69 0.56965 0.69680 0.80770 0.83213 0.88401 1.00000 0.89505 0.88577 0.08523
## 70H79 0.55230 0.69040 0.76756 0.81174 0.87045 0.89505 1.00000 0.88095 0.12376
## 80H   0.60440 0.73100 0.75979 0.78068 0.83564 0.88577 0.88095 1.00000 0.10714
## H?A   0.10335 0.09188 0.07309 0.03805 0.04966 0.08523 0.12376 0.10714 1.00000
```

**Figure 4e**

```
DA<-C[,c("D17","18D29","30D39","40D49","50D59","60D69","70D79","80D","D?A")]
round(cor(DA),5)
```

```
## Warning in cor(DA): the standard deviation is zero
```

```
##            D17   18D29   30D39   40D49    50D59    60D69    70D79     80D D?A
## D17    1.00000 0.06035 0.07340 0.00629 -0.02706 -0.00944 -0.01035 0.02613  NA
## 18D29  0.06035 1.00000 0.03388 0.14008  0.12412  0.13981  0.20999 0.20730  NA
## 30D39  0.07340 0.03388 1.00000 0.17775  0.26228  0.32159  0.28462 0.33226  NA
## 40D49  0.00629 0.14008 0.17775 1.00000  0.40505  0.42029  0.45075 0.46180  NA
## 50D59 -0.02706 0.12412 0.26228 0.40505  1.00000  0.58965  0.59014 0.60102  NA
## 60D69 -0.00944 0.13981 0.32159 0.42029  0.58965  1.00000  0.72457 0.70931  NA
## 70D79 -0.01035 0.20999 0.28462 0.45075  0.59014  0.72457  1.00000 0.76947  NA
## 80D    0.02613 0.20730 0.33226 0.46180  0.60102  0.70931  0.76947 1.00000  NA
## D?A         NA      NA      NA      NA       NA       NA       NA      NA   1
```

**Figure 4f**

```
round(cov(CA),1)
```

```
##           C17   18C29   30C39   40C49   50C59   60C69  70C79    80C   C?A
## C17   57462.0 60101.1 51475.0 35966.2 28571.6 18679.7 8551.9 4890.2 107.0
## 18C29 60101.1 79172.7 66381.7 45411.5 34982.4 22297.2 9916.6 5614.2  97.1
## 30C39 51475.0 66381.7 56266.8 38596.5 29854.0 19109.2 8548.1 4872.9  86.0
## 40C49 35966.2 45411.5 38596.5 27384.0 21503.6 13911.6 6314.2 3652.2  60.4
## 50C59 28571.6 34982.4 29854.0 21503.6 17348.4 11295.7 5170.3 3040.6  48.9
## 60C69 18679.7 22297.2 19109.2 13911.6 11295.7  7526.3 3466.5 2061.7  32.5
## 70C79  8551.9  9916.6  8548.1  6314.2  5170.3  3466.5 1664.0  991.3  14.2
## 80C    4890.2  5614.2  4872.9  3652.2  3040.6  2061.7  991.3  645.2   8.6
## C?A     107.0    97.1    86.0    60.4    48.9    32.5   14.2    8.6   0.8
```

**Figure 4g**

```
round(cov(HA),4)
```

```
##           H17   18H29   30H39   40H49   50H59    60H69   70H79     80H    H?A
## H17    4.6172  6.8697  6.4221  6.2902 10.1052  13.4588 11.0241 11.5976 0.0209
## 18H29  6.8697 19.5335 15.3351 16.4579 26.8347  33.8616 28.3445 28.8509 0.0382
## 30H39  6.4221 15.3351 22.7085 20.4928 34.2026  42.3209 33.9769 32.3326 0.0328
## 40H49  6.2902 16.4579 20.4928 34.4304 45.4983  53.6871 44.2453 40.9068 0.0210
## 50H59 10.1052 26.8347 34.2026 45.4983 86.0178  90.1490 74.9924 69.2098 0.0433
## 60H69 13.4588 33.8616 42.3209 53.6871 90.1490 120.8978 91.4185 86.9725 0.0882
## 70H79 11.0241 28.3445 33.9769 44.2453 74.9924  91.4185 86.2896 73.0777 0.1081
## 80H   11.5976 28.8509 32.3326 40.9068 69.2098  86.9725 73.0777 79.7450 0.0900
## H?A    0.0209  0.0382  0.0328  0.0210  0.0433   0.0882  0.1081  0.0900 0.0088
```

**Figure 4h**

```
round(cov(DA),5)
```

```
##            D17   18D29   30D39   40D49    50D59    60D69    70D79      80D D?A
## D17    0.01010 0.00173 0.00396 0.00057 -0.00407 -0.00255 -0.00312  0.01107   0
## 18D29  0.00173 0.08156 0.00519 0.03588  0.05307  0.10732  0.17969  0.24945   0
## 30D39  0.00396 0.00519 0.28817 0.08557  0.21081  0.46400  0.45780  0.75153   0
## 40D49  0.00057 0.03588 0.08557 0.80417  0.54385  1.01301  1.21114  1.74487   0
## 50D59 -0.00407 0.05307 0.21081 0.54385  2.24179  2.37295  2.64748  3.79161   0
## 60D69 -0.00255 0.10732 0.46400 1.01301  2.37295  7.22419  5.83521  8.03287   0
## 70D79 -0.00312 0.17969 0.45780 1.21114  2.64748  5.83521  8.97764  9.71432   0
## 80D    0.01107 0.24945 0.75153 1.74487  3.79161  8.03287  9.71432 17.75315   0
## D?A    0.00000 0.00000 0.00000 0.00000  0.00000  0.00000  0.00000  0.00000   0
```

**Figure 4j**

```
G<-C[,c("CF","CM","C?G","HF","HM","H?G","DF","DM","D?G")]
round(cor(G),5)
```

```
## Warning in cor(G): the standard deviation is zero
```

```
##          CF      CM     C?G      HF      HM     H?G      DF      DM D?G
## CF  1.00000 0.99672 0.89012 0.77036 0.66910 0.08466 0.33781 0.30049  NA
## CM  0.99672 1.00000 0.87243 0.78988 0.69491 0.08566 0.36470 0.32955  NA
## C?G 0.89012 0.87243 1.00000 0.63695 0.52811 0.06468 0.24144 0.20265  NA
## HF  0.77036 0.78988 0.63695 1.00000 0.94627 0.11475 0.66000 0.66620  NA
## HM  0.66910 0.69491 0.52811 0.94627 1.00000 0.09372 0.69829 0.72448  NA
## H?G 0.08466 0.08566 0.06468 0.11475 0.09372 1.00000 0.09903 0.06348  NA
## DF  0.33781 0.36470 0.24144 0.66000 0.69829 0.09903 1.00000 0.82844  NA
## DM  0.30049 0.32955 0.20265 0.66620 0.72448 0.06348 0.82844 1.00000  NA
## D?G      NA      NA      NA      NA      NA      NA      NA      NA   1
```

**Figure 4k**

```
round(cov(G),2)
```

```
##            CF        CM     C?G       HF       HM  H?G      DF      DM D?G
## CF  410009.28 336241.69 5175.71 12215.50 11330.37 4.72 1025.67 1256.29   0
## CM  336241.69 277561.23 4173.81 10305.25  9682.03 3.93  911.08 1133.59   0
```

```
## C?G   5175.71   4173.81   82.46   143.23   126.83 0.05   10.40   12.02   0
## HF   12215.50  10305.25  143.23   613.25   619.72 0.25   77.50  107.72   0
## HM   11330.37   9682.03  126.83   619.72   699.39 0.22   87.57  125.10   0
## H?G      4.72      3.93    0.05     0.25     0.22 0.01    0.04    0.04   0
## DF    1025.67    911.08   10.40    77.50    87.57 0.04   22.48   25.65   0
## DM    1256.29   1133.59   12.02   107.72   125.10 0.04   25.65   42.63   0
## D?G      0.00      0.00    0.00     0.00     0.00 0.00    0.00    0.00   0
```

**Figure 4m**

```r
CR<-C[,c("CW","CB","CL","CA","Co","C?R")]
cor(CR)
```

```
##            CW        CB        CL        CA        Co       C?R
## CW  1.0000000 0.9491987 0.8543795 0.9443006 0.9588982 0.9456685
## CB  0.9491987 1.0000000 0.8318563 0.9037324 0.9648302 0.9480892
## CL  0.8543795 0.8318563 1.0000000 0.8547730 0.7921702 0.8412106
## CA  0.9443006 0.9037324 0.8547730 1.0000000 0.9201286 0.9614993
## Co  0.9588982 0.9648302 0.7921702 0.9201286 1.0000000 0.9567045
## C?R 0.9456685 0.9480892 0.8412106 0.9614993 0.9567045 1.0000000
```

**Figure 4n**

```r
HR<-C[,c("HW","HB","HL","HA","Ho","H?R")]
cor(HR)
```

```
##            HW        HB        HL        HA        Ho       H?R
## HW  1.0000000 0.8470814 0.7952327 0.6784519 0.7429546 0.5511925
## HB  0.8470814 1.0000000 0.7719411 0.6736161 0.7402507 0.5103701
## HL  0.7952327 0.7719411 1.0000000 0.6924339 0.7620071 0.4832228
## HA  0.6784519 0.6736161 0.6924339 1.0000000 0.6191852 0.3940523
## Ho  0.7429546 0.7402507 0.7620071 0.6191852 1.0000000 0.3391114
## H?R 0.5511925 0.5103701 0.4832228 0.3940523 0.3391114 1.0000000
```

**Figure 4p**

```r
DR<-C[,c("DW","DB","DL","DA","Do","D?R")]
cor(DR)
```

```
##            DW         DB         DL        DA        Do        D?R
## DW  1.0000000 0.69908125 0.67873318 0.4833451 0.2614516 0.14828577
## DB  0.6990813 1.00000000 0.66727406 0.5622825 0.2647635 0.09105115
## DL  0.6787332 0.66727406 1.00000000 0.5632928 0.2858164 0.09669581
## DA  0.4833451 0.56228249 0.56329284 1.0000000 0.1378129 0.14510074
## Do  0.2614516 0.26476351 0.28581635 0.1378129 1.0000000 0.03405010
## D?R 0.1482858 0.09105115 0.09669581 0.1451007 0.0340501 1.00000000
```

**Figure 4q**

```r
cov(CR)
```

```
##          CW       CB       CL        CA       Co      C?R
## CW  77430.19 72902.46 71912.03 12355.897 22807.78 62129.43
## CB  72902.46 76183.28 69450.24 11729.474 22763.34 61784.89
## CL  71912.03 69450.24 91493.69 12157.807 20481.86 60076.35
```

```
## CA  12355.90 11729.47 12157.81  2211.148  3698.39 10674.83
## Co  22807.78 22763.34 20481.86  3698.390  7306.52 19307.96
## C?R 62129.43 61784.89 60076.35 10674.833 19307.96 55745.08
```

### Figure 4r

```
cov(HR)
```

```
##             HW        HB         HL         HA         Ho        H?R
## HW  100.170714 214.05695 115.697145 12.9569966 16.6777606  6.7878591
## HB  214.056946 637.48147 283.319183 32.4534674 41.9196559 15.8554368
## HL  115.697145 283.31918 211.308079 19.2066343 24.8440368  8.6430034
## HA   12.956997  32.45347  19.206634  3.6410763  2.6499675  0.9251836
## Ho   16.677761  41.91966  24.844037  2.6499675  5.0304888  0.9358508
## H?R   6.787859  15.85544   8.643003  0.9251836  0.9358508  1.5139737
```

### Figure 4s

```
cov(DR)
```

```
##             DW          DB          DL         DA          Do         D?R
## DW  7.33597101  9.29595086  6.99050273 1.06046893 0.204130378 0.053189588
## DB  9.29595086 24.10320096 12.45724847 2.23616437 0.374699402 0.059199922
## DL  6.99050273 12.45724847 14.45975887 1.73510822 0.313296178 0.048695242
## DA  1.06046893  2.23616437  1.73510822 0.65618094 0.032180229 0.015566099
## Do  0.20413038  0.37469940  0.31329618 0.03218023 0.083095021 0.001299883
## D?R 0.05318959  0.05919992  0.04869524 0.01556610 0.001299883 0.017538672
```

### Figure 5a

```
prcomp(C[,c("C","H","D")])
```

```
## Standard deviations (1, .., p=3):
## [1] 1174.837195   34.928538    6.508861
##
## Rotation (n x k) = (3 x 3):
##           PC1         PC2          PC3
## C 0.999489716  0.03167598  0.004115843
## H 0.031785689 -0.97355736 -0.226220560
## D 0.003158748 -0.22623595  0.974067409
```

```
prcomp(C[,c("C","H","D")])$sdev[1]^2/sum(prcomp(C[,c("C","H","D")])$sdev^2)
```

```
## [1] 0.9990862
```

### Figure 5b

```
prcomp(CA)
```

```
## Standard deviations (1, .., p=9):
## [1] 485.3907826  91.2560747  53.7281763  17.0536181  13.0573057  10.3810008
## [7]   7.2929323   5.3678585   0.7226089
##
## Rotation (n x k) = (9 x 9):
##                 PC1          PC2           PC3         PC4           PC5
```

```
## C17   0.4662744602  0.855847545  0.2191196489 -0.02528295  0.0352204040
## 18C29 0.5723882923 -0.423105383  0.3666288396 -0.59416218 -0.0215320640
## 30C39 0.4850510124 -0.266363853  0.1501489394  0.74254567 -0.3100155279
## 40C49 0.3369534696 -0.114380739 -0.3763264324  0.19941847  0.6997735738
## 50C59 0.2635519455 -0.003696933 -0.5630884744 -0.16674798  0.0816421692
## 60C69 0.1699521012  0.043278771 -0.4704095372 -0.15331647 -0.4192107719
## 70C79 0.0766557418  0.042449432 -0.2694081321 -0.05356876 -0.3087985206
## 80C   0.0439040419  0.027892796 -0.2060800740 -0.03145580 -0.3670737173
## C?A   0.0007955287  0.002734658  0.0002905727  0.00451277  0.0006751478
##                PC6          PC7           PC8           PC9
## C17   -0.006402688  0.007213856  0.0109054749  2.831903e-03
## 18C29 -0.069129282  0.025502897  0.0007795336 -3.819115e-03
## 30C39  0.150063913 -0.018042292 -0.0298825428  3.119949e-03
## 40C49 -0.441901279 -0.013589740  0.0827346362  7.735569e-04
## 50C59  0.717059704  0.247965071 -0.0576753342  5.989103e-05
## 60C69 -0.163044036 -0.705418031  0.1564341468  6.699972e-03
## 70C79 -0.387437448  0.317811348 -0.7549666690 -2.070176e-02
## 80C   -0.294094165  0.581919354  0.6276292143  7.273628e-03
## C?A    0.005205752 -0.007203806  0.0212435431 -9.997203e-01
```

```r
prcomp(CA)$sdev[1]^2/sum(prcomp(CA)$sdev^2)
```

```
## [1] 0.9520508
```

**Figure 5c**

```r
prcomp(HA)
```

```
## Standard deviations (1, .., p=9):
## [1] 19.82764324  3.94647474  3.36934033  3.27892712  2.99492915  2.76836206
## [7]  2.21989520  1.36556381  0.09233181
##
## Rotation (n x k) = (9 x 9):
##                PC1          PC2          PC3           PC4          PC5
## H17   0.0658216625  0.183952645 -0.193646255  0.1332904464 -0.078695446
## 18H29 0.1676775720  0.379152594 -0.444791470  0.4113538898 -0.203580555
## 30H39 0.2021972054  0.075564999 -0.466079973  0.1584198132 -0.252109192
## 40H49 0.2597468772 -0.272671856 -0.165707462  0.1269842774 -0.128860977
## 50H59 0.4396006857 -0.642858185 -0.010619707  0.3909643547  0.344525162
## 60H69 0.5371014305 -0.020018503 -0.277108964 -0.7698846872 -0.008178566
## 70H79 0.4458310466  0.073735079  0.621077927  0.1043670055 -0.624186740
## 80H   0.4227379404  0.568535193  0.239432105  0.1238213241  0.603174551
## H?A   0.0004388174  0.002865412  0.002119091 -0.0001996765 -0.002164821
##                PC6           PC7          PC8           PC9
## H17   -0.046348044  8.709906e-02 -0.943731677 -2.623001e-03
## 18H29 -0.101444275  5.604258e-01  0.308646984 -3.104227e-05
## 30H39 -0.174429274 -7.785615e-01  0.104575861  2.283311e-04
## 40H49  0.893146607  5.547787e-03 -0.015708127  2.028155e-03
## 50H59 -0.333269208  9.530432e-02 -0.040819017  1.904486e-03
## 60H69 -0.106461504  1.737682e-01  0.003630304  7.677736e-05
## 70H79 -0.096963138 -5.246054e-03 -0.010896063 -3.220419e-03
## 80H    0.166186928 -1.812235e-01  0.033479562 -6.543513e-04
## H?A   -0.001456987  8.205132e-05 -0.002393600  9.999873e-01
```

```
prcomp(HA)$sdev[1]^2/sum(prcomp(HA)$sdev^2)
```

```
## [1] 0.8654632
```

**Figure 5d**

```
DAa<-DA[,-9]
prcomp(DAa)
```

```
## Standard deviations (1, .., p=8):
## [1] 5.46307244 1.78657690 1.47593995 1.11257505 0.76773947 0.50109775 0.27760045
## [8] 0.09971939
##
## Rotation (n x k) = (8 x 8):
##                 PC1          PC2           PC3          PC4          PC5
## D17   -0.0001694227 -0.003335410 -0.0008073472 -0.003109129 -0.001710687
## 18D29 -0.0110805773 -0.004432816  0.0155532779 -0.002794046 -0.018013416
## 30D39 -0.0344400812  0.004274232 -0.0310504772  0.016344064  0.003347132
## 40D49 -0.0828570274  0.023590949  0.0065422932  0.139126054 -0.986326544
## 50D59 -0.1859637910  0.115418550 -0.0695607014  0.960755796  0.153387849
## 60D69 -0.4129077287  0.604488583 -0.6491620763 -0.203046797  0.016108984
## 70D79 -0.4891618161  0.435108966  0.7479612821 -0.099388587  0.042700309
## 80D   -0.7399129541 -0.656769916 -0.1142555823 -0.078749901  0.034794942
##                PC6          PC7           PC8
## D17   -0.015756012  0.025902918  0.9995280828
## 18D29  0.028662295  0.998905746 -0.0254786022
## 30D39 -0.998235283  0.028450383 -0.0164330156
## 40D49  0.001214777 -0.018366172 -0.0006902536
## 50D59  0.025379560  0.004355432  0.0038356868
## 60D69  0.033935848  0.006988630  0.0011726418
## 70D79 -0.006453966 -0.014412807  0.0020088840
## 80D    0.024852527 -0.009698878 -0.0019516343
```

```
prcomp(DAa)$sdev[1]^2/sum(prcomp(DAa)$sdev^2)
```

```
## [1] 0.7984095
```

**Figure 5e**

```
prcomp(CR)
```

```
## Standard deviations (1, .., p=6):
## [1] 533.41990 135.16066  62.08675  57.82069  15.99341  10.53258
##
## Rotation (n x k) = (6 x 6):
##            PC1         PC2          PC3          PC4         PC5         PC6
## CW  0.50909550 -0.25418359 -0.772020170  0.246196828  0.11246911 -0.08327759
## CB  0.50150873 -0.36497000  0.613805902  0.466809436  0.13425255  0.05092550
## CL  0.52486044  0.84645852  0.068801456  0.023445241 -0.05093873 -0.01230561
## CA  0.08391405 -0.02201713 -0.074736310 -0.143298732  0.09551395  0.97838169
## Co  0.15278829 -0.15179150 -0.002110603 -0.008499842 -0.97344395  0.07710551
## C?R 0.42830650 -0.24933946  0.130013698 -0.836851014  0.10008080 -0.16475444
```

```
prcomp(CR)$sdev[1]^2/sum(prcomp(CR)$sdev^2)
```

```
## [1] 0.9167667
```

**Figure 5f**

```r
prcomp(HR)
```

```
## Standard deviations (1, .., p=6):
## [1] 29.3670660  8.4451426  4.5745192  1.3817465  1.2735564  0.9712903
##
## Rotation (n x k) = (6 x 6):
##            PC1         PC2         PC3         PC4          PC5         PC6
## HW  0.30511807  0.14275790  0.93819882 -0.04631811  0.024354596 -0.05972939
## HB  0.84756040 -0.48944563 -0.20327824 -0.02406095  0.006577802 -0.01183641
## HL  0.42696610  0.85792891 -0.27498433 -0.07140912  0.017577564 -0.02516577
## HA  0.04639095  0.03818364  0.02207824  0.44258871 -0.893841353 -0.03263906
## Ho  0.05991364  0.04960859  0.02463791  0.84724991  0.413552866  0.32323976
## H?R 0.02241673  0.01010626  0.04183387 -0.28010582 -0.170517675  0.94345627
```

```r
prcomp(HR)$sdev[1]^2/sum(prcomp(HR)$sdev^2)
```

```
## [1] 0.899159
```

**Figure 5g**

```r
prcomp(DR)
```

```
## Standard deviations (1, .., p=6):
## [1] 6.1173827 2.4490660 1.6552439 0.6354768 0.2735398 0.1302421
##
## Rotation (n x k) = (6 x 6):
##             PC1          PC2          PC3          PC4          PC5
## DW  0.361714841  0.134524779  0.922403761 -0.009642552 -0.009102617
## DB  0.762399094 -0.609564969 -0.210675346 -0.052288156 -0.007523611
## DL  0.530101061  0.780794211 -0.322626098 -0.070967120 -0.015071249
## DA  0.081833692  0.025197970 -0.025199840  0.995151188  0.035130753
## Do  0.014146571  0.007521730  0.002812482 -0.036609639  0.999194229
## D?R 0.002445648  0.001587204  0.007527236  0.021690674  0.003180999
##              PC6
## DW  -0.007805318
## DB   0.001847350
## DL   0.001480433
## DA  -0.021753633
## Do  -0.002452723
## D?R  0.999727081
```

```r
prcomp(DR)$sdev[1]^2/sum(prcomp(DR)$sdev^2)
```

```
## [1] 0.8020957
```

**Figure 6a**

```r
factanal(CA,5)
```

```
##
## Call:
## factanal(x = CA, factors = 5)
```

```
##
## Uniquenesses:
##   C17 18C29 30C39 40C49 50C59 60C69 70C79   80C   C?A
## 0.005 0.005 0.005 0.005 0.005 0.007 0.005 0.065 0.693
##
## Loadings:
##       Factor1 Factor2 Factor3 Factor4 Factor5
## C17    0.454   0.533   0.692   0.159
## 18C29  0.432   0.820   0.367
## 30C39  0.462   0.795   0.385
## 40C49  0.577   0.717   0.380
## 50C59  0.657   0.626   0.405
## 60C69  0.719   0.553   0.409
## 70C79  0.799   0.469   0.361
## 80C    0.830   0.375   0.325
## C?A    0.179   0.174   0.494
##
##                Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings      3.247   3.193   1.718   0.036   0.012
## Proportion Var   0.361   0.355   0.191   0.004   0.001
## Cumulative Var   0.361   0.716   0.907   0.911   0.912
##
## Test of the hypothesis that 5 factors are sufficient.
## The chi square statistic is 51.87 on 1 degree of freedom.
## The p-value is 5.93e-13
```

```
round(factanal(CA,5)$loadings%*%t(factanal(CA,5)$loadings)+factanal(CA,5)$uniqueness,4)
cor(CA)
factanal(CA,5,scores="Bartlett")$scores
factanal(CA,5,scores="regression")$scores
```

**Figure 6b**

```
factanal(CA,4)
```

```
##
## Call:
## factanal(x = CA, factors = 4)
##
## Uniquenesses:
##   C17 18C29 30C39 40C49 50C59 60C69 70C79   80C   C?A
## 0.005 0.005 0.005 0.006 0.005 0.009 0.011 0.062 0.703
##
## Loadings:
##       Factor1 Factor2 Factor3 Factor4
## C17    0.451   0.523   0.711   0.113
## 18C29  0.431   0.817   0.375
## 30C39  0.462   0.791   0.395
## 40C49  0.576   0.710   0.395
## 50C59  0.655   0.621   0.414
## 60C69  0.718   0.551   0.411
## 70C79  0.793   0.463   0.382
## 80C    0.833   0.373   0.322
## C?A    0.180   0.174   0.484
```

```
##
##                 Factor1 Factor2 Factor3 Factor4
## SS loadings       3.235   3.144   1.782   0.029
## Proportion Var    0.359   0.349   0.198   0.003
## Cumulative Var    0.359   0.709   0.907   0.910
##
## Test of the hypothesis that 4 factors are sufficient.
## The chi square statistic is 91.76 on 6 degrees of freedom.
## The p-value is 1.31e-17
```

```
round(factanal(CA,4)$loadings%*%t(factanal(CA,4)$loadings)+factanal(CA,4)$uniqueness,4)
cor(CA)
factanal(CA,4,scores="Bartlett")$scores
factanal(CA,4,scores="regression")$scores
```

**Figure 6c**

```
factanal(HA,5)
```

```
##
## Call:
## factanal(x = HA, factors = 5)
##
## Uniquenesses:
##   H17 18H29 30H39 40H49 50H59 60H69 70H79   80H   H?A
## 0.313 0.208 0.234 0.220 0.100 0.005 0.118 0.082 0.005
##
## Loadings:
##       Factor1 Factor2 Factor3 Factor4 Factor5
## H17    0.307   0.767
## 18H29  0.475   0.750
## 30H39  0.674   0.547
## 40H49  0.817   0.328
## 50H59  0.893   0.307
## 60H69  0.884   0.373                   0.269
## 70H79  0.852   0.371           0.113
## 80H    0.793   0.457           0.276
## H?A                    0.996
##
##                 Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings       4.378   2.140   1.008   0.106   0.082
## Proportion Var    0.486   0.238   0.112   0.012   0.009
## Cumulative Var    0.486   0.724   0.836   0.848   0.857
##
## Test of the hypothesis that 5 factors are sufficient.
## The chi square statistic is 0.18 on 1 degree of freedom.
## The p-value is 0.67
```

```
round(factanal(HA,5)$loadings%*%t(factanal(HA,5)$loadings)+factanal(HA,5)$uniqueness,4)
cor(HA)
factanal(HA,4,scores="Bartlett")$scores
factanal(HA,4,scores="regression")$scores
```

**Figure 6d**

```
factanal(HA,2)
```

```
##
## Call:
## factanal(x = HA, factors = 2)
##
## Uniquenesses:
##   H17 18H29 30H39 40H49 50H59 60H69 70H79   80H   H?A
## 0.298 0.217 0.258 0.236 0.132 0.087 0.119 0.145 0.986
##
## Loadings:
##       Factor1 Factor2
## H17   0.283   0.789
## 18H29 0.458   0.757
## 30H39 0.662   0.551
## 40H49 0.804   0.343
## 50H59 0.869   0.335
## 60H69 0.863   0.410
## 70H79 0.850   0.399
## 80H   0.789   0.482
## H?A           0.107
##
##                Factor1 Factor2
## SS loadings      4.222   2.299
## Proportion Var   0.469   0.255
## Cumulative Var   0.469   0.725
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 100.62 on 19 degrees of freedom.
## The p-value is 4.14e-13
```

```
round(factanal(HA,2)$loadings%*%t(factanal(HA,2)$loadings)+factanal(HA,2)$uniqueness,4)
cor(HA)
factanal(HA,2,scores="Bartlett")$scores
factanal(HA,2,scores="regression")$scores
```

**Figure 6e**

```
factanal(DAa,4)
```

```
##
## Call:
## factanal(x = DAa, factors = 4)
##
## Uniquenesses:
##   D17 18D29 30D39 40D49 50D59 60D69 70D79   80D
## 0.903 0.836 0.720 0.687 0.433 0.309 0.182 0.240
##
## Loadings:
##       Factor1 Factor2 Factor3 Factor4
## D17                   0.297
## 18D29  0.126   0.373
## 30D39  0.398  -0.124   0.325
```

```
## 40D49  0.500   0.211           -0.131
## 50D59  0.721                   -0.191
## 60D69  0.823
## 70D79  0.832   0.290            0.201
## 80D    0.826   0.251   0.104
##
##                Factor1 Factor2 Factor3 Factor4
## SS loadings      2.998   0.371   0.217   0.103
## Proportion Var   0.375   0.046   0.027   0.013
## Cumulative Var   0.375   0.421   0.448   0.461
##
## Test of the hypothesis that 4 factors are sufficient.
## The chi square statistic is 0.14 on 2 degrees of freedom.
## The p-value is 0.934
```

```
round(factanal(DAa,4)$loadings%*%t(factanal(DAa,4)$loadings)+factanal(DAa,4)$uniqueness,5)
cor(DAa)
factanal(DAa,4,scores="Bartlett")$scores
factanal(DAa,4,scores="regression")$scores
```

**Figure 6f**

```
factanal(DAa,1)
```

```
##
## Call:
## factanal(x = DAa, factors = 1)
##
## Uniquenesses:
##   D17 18D29 30D39 40D49 50D59 60D69 70D79   80D
## 1.000 0.953 0.869 0.724 0.519 0.323 0.237 0.236
##
## Loadings:
##       Factor1
## D17
## 18D29 0.218
## 30D39 0.362
## 40D49 0.526
## 50D59 0.693
## 60D69 0.823
## 70D79 0.874
## 80D   0.874
##
##                Factor1
## SS loadings      3.140
## Proportion Var   0.393
##
## Test of the hypothesis that 1 factor is sufficient.
## The chi square statistic is 37.39 on 20 degrees of freedom.
## The p-value is 0.0105
```

```
round(factanal(DAa,1)$loadings%*%t(factanal(DAa,1)$loadings)+factanal(DAa,1)$uniqueness,5)
cor(DAa)
factanal(DAa,1,scores="Bartlett")$scores
factanal(DAa,1,scores="regression")$scores
```

### Figure 6g

```
factanal(CR,3)
```

```
##
## Call:
## factanal(x = CR, factors = 3)
##
## Uniquenesses:
##    CW    CB    CL    CA    Co   C?R
## 0.045 0.040 0.005 0.005 0.016 0.036
##
## Loadings:
##     Factor1 Factor2 Factor3
## CW  0.734   0.493   0.415
## CB  0.803   0.473   0.302
## CL  0.421   0.854   0.298
## CA  0.604   0.483   0.629
## Co  0.840   0.389   0.357
## C?R 0.721   0.462   0.480
##
##                Factor1 Factor2 Factor3
## SS loadings      2.953   1.794   1.106
## Proportion Var   0.492   0.299   0.184
## Cumulative Var   0.492   0.791   0.975
##
## The degrees of freedom for the model is 0 and the fit was 0.1183
```

```
factanal(CR,3)$loadings%*%t(factanal(CR,3)$loadings)+factanal(CR,3)$uniqueness
cor(CR)
factanal(CR,3,scores="Bartlett")$scores
factanal(CR,3,scores="regression")$scores
```

### Figure 6h

```
factanal(CR,2)
```

```
##
## Call:
## factanal(x = CR, factors = 2)
##
## Uniquenesses:
##    CW    CB    CL    CA    Co   C?R
## 0.048 0.033 0.257 0.005 0.031 0.036
##
## Loadings:
##     Factor1 Factor2
## CW  0.713   0.666
## CB  0.811   0.556
## CL  0.541   0.671
## CA  0.540   0.838
## Co  0.789   0.589
```

```
## C?R 0.682   0.707
##
##                 Factor1 Factor2
## SS loadings      2.838   2.752
## Proportion Var   0.473   0.459
## Cumulative Var   0.473   0.932
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 272.23 on 4 degrees of freedom.
## The p-value is 1.05e-57
```

```
factanal(CR,2)$loadings%*%t(factanal(CR,2)$loadings)+factanal(CR,2)$uniqueness
cor(CR)
factanal(CR,2,scores="Bartlett")$scores
factanal(CR,2,scores="regression")$scores
```

**Figure 6j**

```
factanal(HR,3)
```

```
##
## Call:
## factanal(x = HR, factors = 3)
##
## Uniquenesses:
##    HW    HB    HL    HA    Ho   H?R
## 0.148 0.153 0.050 0.440 0.303 0.005
##
## Loadings:
##     Factor1 Factor2 Factor3
## HW  0.831   0.335   0.221
## HB  0.855   0.292   0.176
## HL  0.687   0.243   0.648
## HA  0.656   0.205   0.296
## Ho  0.759   0.118   0.327
## H?R 0.243   0.959   0.129
##
##                 Factor1 Factor2 Factor3
## SS loadings      2.959   1.232   0.710
## Proportion Var   0.493   0.205   0.118
## Cumulative Var   0.493   0.698   0.817
##
## The degrees of freedom for the model is 0 and the fit was 0
```

```
factanal(HR,3)$loadings%*%t(factanal(HR,3)$loadings)+factanal(HR,3)$uniqueness
cor(HR)
factanal(HR,3,scores="Bartlett")$scores
factanal(HR,3,scores="regression")$scores
```

**Figure 6k**

```
factanal(HR,2)
```

```
##
## Call:
```

```
## factanal(x = HR, factors = 2)
##
## Uniquenesses:
##    HW    HB    HL    HA    Ho   H?R
## 0.131 0.184 0.234 0.444 0.146 0.580
##
## Loadings:
##     Factor1 Factor2
## HW  0.632   0.685
## HB  0.646   0.632
## HL  0.704   0.520
## HA  0.568   0.482
## Ho  0.885   0.265
## H?R 0.202   0.616
##
##                Factor1 Factor2
## SS loadings       2.46   1.820
## Proportion Var    0.41   0.303
## Cumulative Var    0.41   0.713
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 25.35 on 4 degrees of freedom.
## The p-value is 4.28e-05
```

```
factanal(HR,2)$loadings%*%t(factanal(HR,2)$loadings)+factanal(HR,2)$uniqueness
cor(HR)
factanal(HR,2,scores="Bartlett")$scores
factanal(HR,2,scores="regression")$scores
```

### Figure 6m

```
factanal(DR,3)
```

```
##
## Call:
## factanal(x = DR, factors = 3)
##
## Uniquenesses:
##    DW    DB    DL    DA    Do   D?R
## 0.005 0.349 0.308 0.005 0.865 0.965
##
## Loadings:
##     Factor1 Factor2 Factor3
## DW  0.600   0.130   0.786
## DB  0.661   0.324   0.331
## DL  0.719   0.328   0.260
## DA  0.268   0.926   0.257
## Do  0.363
## D?R         0.107   0.152
##
##                Factor1 Factor2 Factor3
## SS loadings      1.518   1.099   0.887
## Proportion Var   0.253   0.183   0.148
## Cumulative Var   0.253   0.436   0.584
```

```
##
## The degrees of freedom for the model is 0 and the fit was 6e-04
```

```
factanal(DR,3)$loadings%*%t(factanal(DR,3)$loadings)+factanal(DR,3)$uniqueness
cor(DR)
factanal(DR,3,scores="Bartlett")$scores
factanal(DR,3,scores="regression")$scores
```

**Figure 6n**

```
factanal(DR,1)
```

```
##
## Call:
## factanal(x = DR, factors = 1)
##
## Uniquenesses:
##    DW    DB    DL    DA    Do   D?R
## 0.327 0.299 0.326 0.581 0.900 0.980
##
## Loadings:
##     Factor1
## DW  0.820
## DB  0.837
## DL  0.821
## DA  0.648
## Do  0.316
## D?R 0.143
##
##                Factor1
## SS loadings      2.587
## Proportion Var   0.431
##
## Test of the hypothesis that 1 factor is sufficient.
## The chi square statistic is 41.92 on 9 degrees of freedom.
## The p-value is 3.4e-06
```

```
factanal(DR,1)$loadings%*%t(factanal(DR,1)$loadings)+factanal(DR,1)$uniqueness
cor(DR)
factanal(DR,1,scores="Bartlett")$scores
factanal(DR,1,scores="regression")$scores
```

## Works Cited


"COVID-19 Daily Cases, Deaths, and Hospitalizations." *Chicago Data Portal*, 24 April 2020, https://data.cityofchicago.org/d/naz8-j4nc. Accessed 29 April 2022.