LDSS Covid-19 Project

Our goal is to determine and analyze the factors that have impacted us during the ongoing COVID-19 Pandemic. In this project, we look at Mobility & Unemployment factors of COVID-19.

We used datasets from authentic websites to research to the intricate correlations that we discovered. We noticed some interesting factors specifically within the analyzed datasets and so, we wanted to further investigate the question, "In what different ways has COVID-19 left an impact on America?"

Datasets: -Datasets we used: apple mobility trends, unemployment rate, us_counties_covid19_daily -apple_mobility: contains data regarding walking, driving and transit -Unemployment: contains unemployment rates for USA from January 1948 up until November 2020. -us_covid: contains the states & counties daily count of COVID cases from January 1st 2020 to March 10th 2020

Variables: -Quantitative: -apple_mobility: dates of volume% change for transportation type -Unemployment: DATE, UNRATE -us_covid: date,fips,cases,deaths -Qualitative: -apple_mobility: geo type, region, transportation type, alternative name, sub.reigon, country -us_covid: county,state

Importing the Datasets

```
apple_mobility <- read.csv("apple mobility trends.csv")

Unemployement <- read.csv("unemployment rate.csv")

us_covid <- read.csv("us_counties_covid19_daily.csv")
```

Cleaning the Unemployment Rate and US Counties Covid19 Daily data sets

```
#Clean data
#apple_mobility2 <- select(apple_mobility,-c(geo_type,sub.region,country,alternative_name))


Unemployement_before_covid <- slice(Unemployement,1:868)
Unemployement_during_covid <- slice(Unemployement,868:875)
head(us_covid)
```

```
##         date    county      state  fips cases deaths
## 1 2020-01-21 Snohomish Washington 53061     1      0
## 2 2020-01-22 Snohomish Washington 53061     1      0
## 3 2020-01-23 Snohomish Washington 53061     1      0
## 4 2020-01-24      Cook   Illinois 17031     1      0
## 5 2020-01-24 Snohomish Washington 53061     1      0
## 6 2020-01-25    Orange California  6059     1      0
```

```
apple_mobility2 <- select(apple_mobility,-c(sub.region,alternative_name))

#cases from jan 21st 2020 --> march 10th 2020
setDT(us_covid)
```

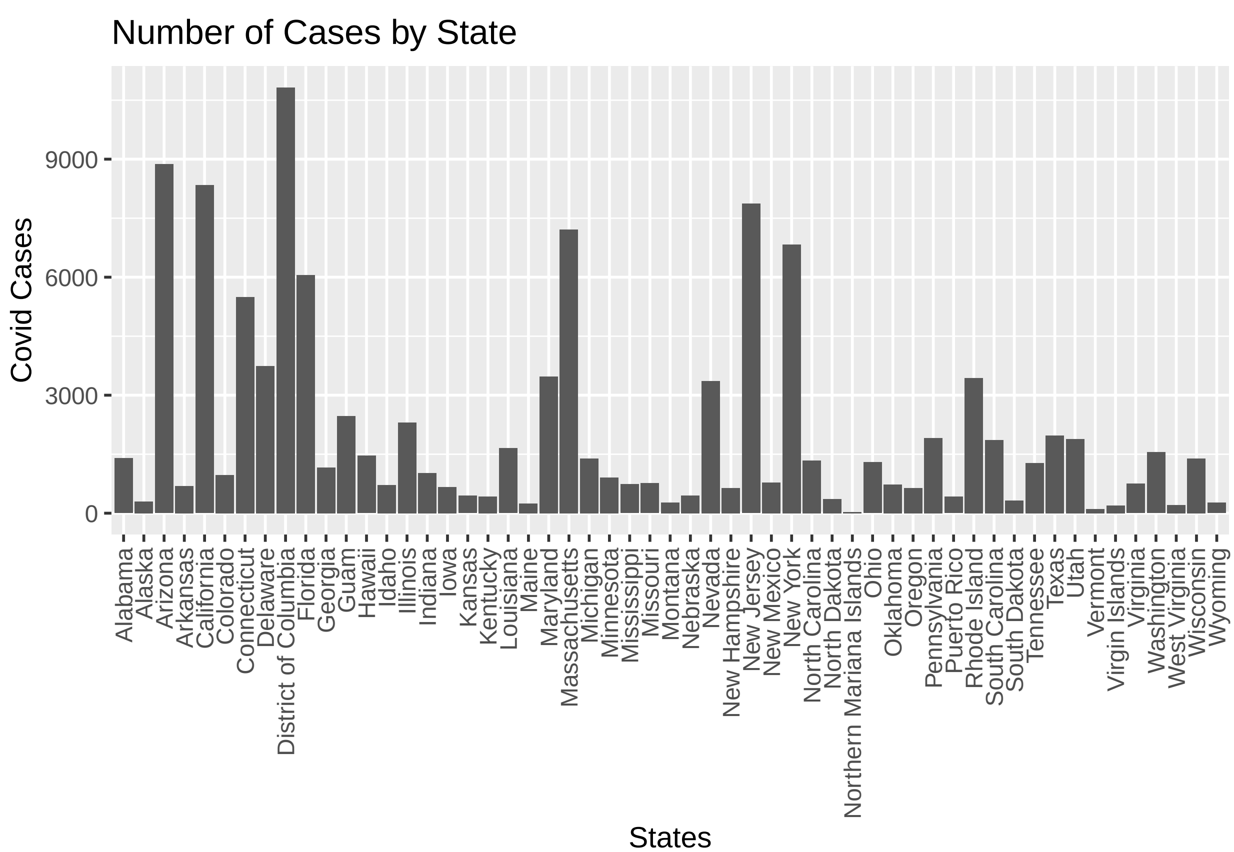
```r
Total_case = us_covid[ , .(cases = mean(cases)), by = .(state)] %>%
  arrange(desc(cases))

Total_case %>%
  mutate(cases=round(Total_case$cases,digits = 0)) %>%
  head(10)
```

```
##                   state cases
##  1: District of Columbia 10823
##  2:              Arizona  8871
##  3:           California  8336
##  4:           New Jersey  7866
##  5:        Massachusetts  7211
##  6:             New York  6831
##  7:              Florida  6053
##  8:          Connecticut  5492
##  9:             Delaware  3745
## 10:             Maryland  3472
```

plotting the Total cases of each state

```r
ggplot(Total_case,aes(x=state,y=cases))+ geom_bar(stat="identity", position=position_dodge(width=100)) +
  xlab("States") + ylab("Covid Cases") + ggtitle("Number of Cases by State")
```

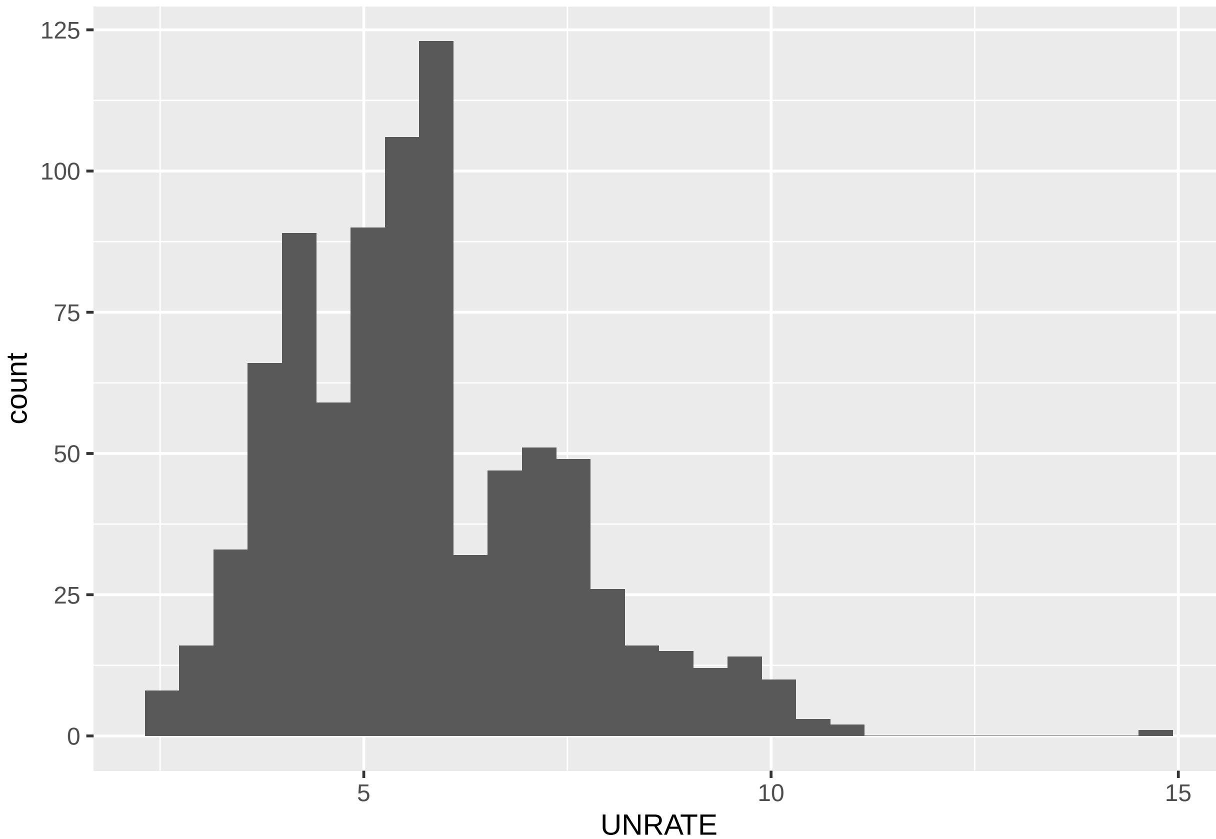


After observing the graph above, we can infer that the highly populated states have the higher cases of

COVID positive people. The climate is also in effect here, although it is said the warmer areas tend to have less Covid going around, the numbers here beg to differ. Places such as Arizona, California, Florida, and District of Columbia, have higher cases compared to Michigan, Missouri, and Vermont for example.

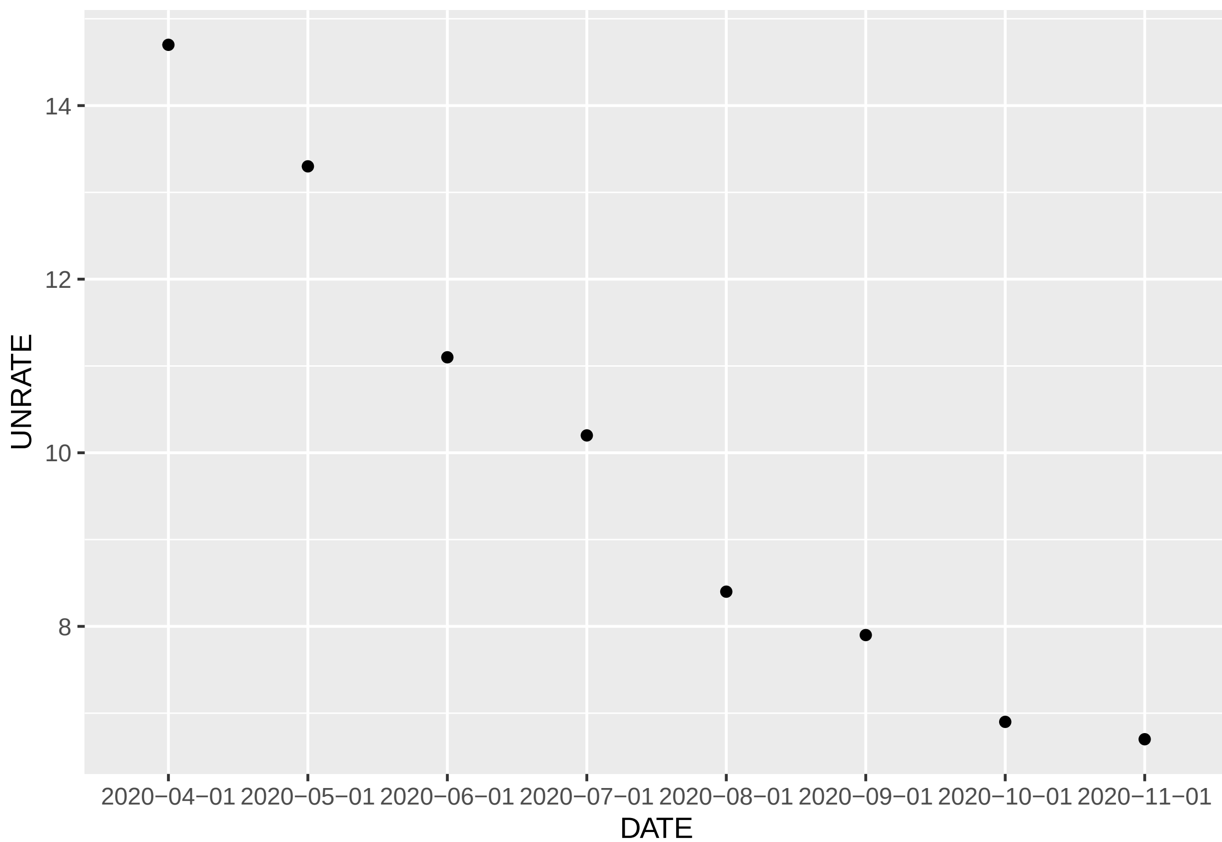
```
ggplot(Unemployement_before_covid,aes(x=UNRATE),binwidth=20) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



In this histogram graph represents the distribution for the unemployment rates from 1948 to march 2020. The graph is clearly skewed to the right which would suggest that the mean is greater than the median and unemployment rate is typically around 5.74%. Lets look at the unemployment rate during covid.

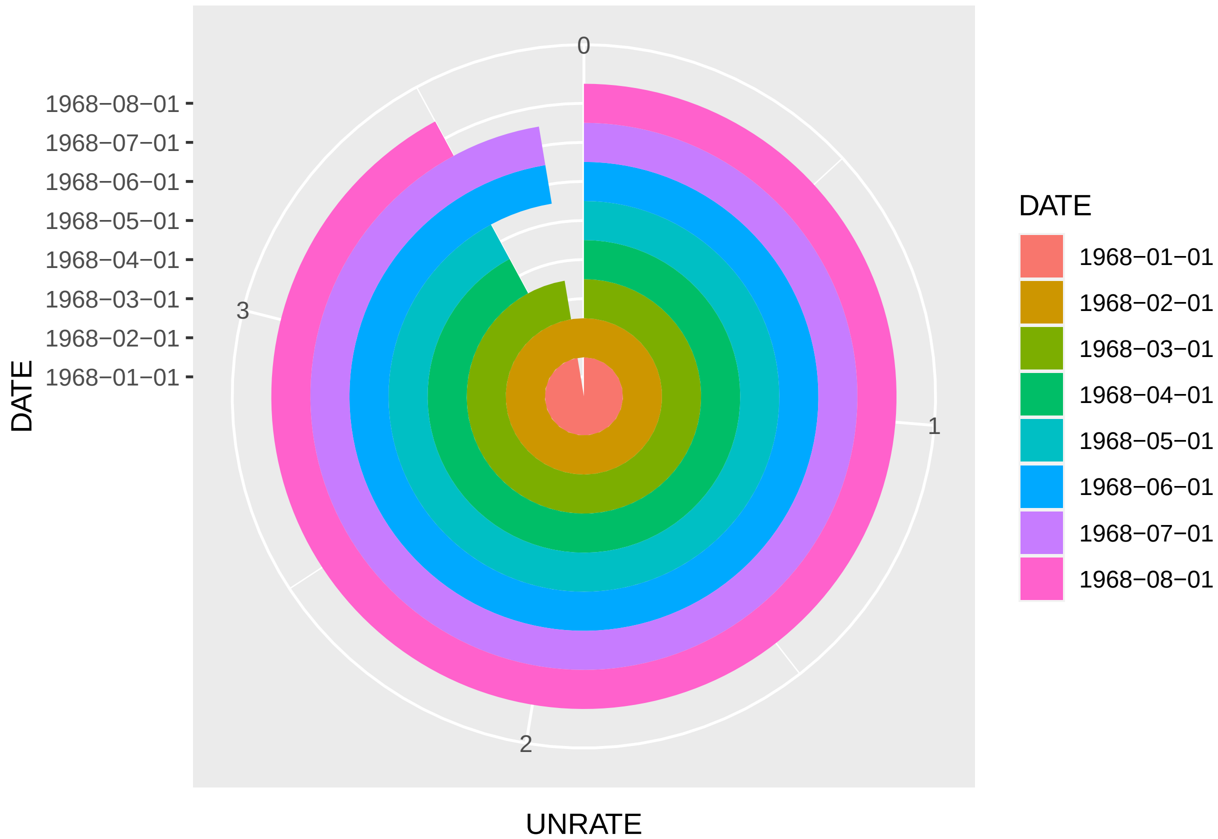
```
ggplot(Unemployement_during_covid,aes(x=DATE,y=UNRATE)) + geom_point()
```



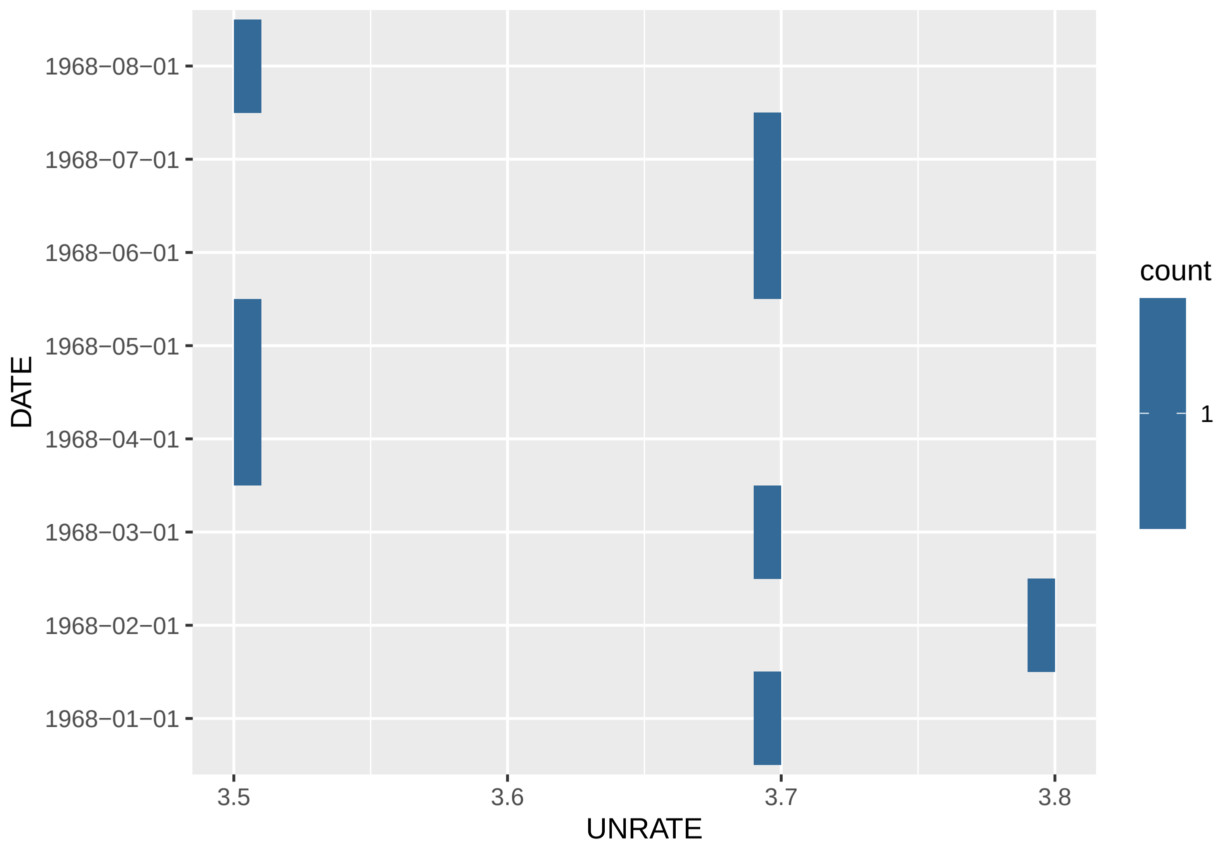
We can see that there is a negative relationship between the unemployment rate and the date. This is because when lockdown started, unemployment rates drastically increased and all of the managers did not know the procedures and protocols of a pandemic. This led to drastic changes in our everyday life and after a few months, notably after April the unemployment rates began decreasing again as most businesses were successful in implementing a way to get their workers to work in a good health and safety environment which would reduce the risk in getting the virus. The pandemic has taught us how to adapt to a pandemic and now we know that if another outbreak of a new virus happens in the future, we will be able to stop it before it gets out of hand unlike this one. According to CIDRAP news, about 20 million jobs were lost due to the pandemic.

Lets compare the Coivd-19 Pandemic to the H3N2 Pandemic which occurred during 1968

```
H3N2_Pandemic <- slice(Unemployement_before_covid,241:248)
bp<- ggplot(H3N2_Pandemic, aes(x=DATE, y=UNRATE, fill=DATE))+
geom_bar(width = 1, stat = "identity")
pie <- bp + coord_polar("y", start=0)
pie
```




```
ggplot(H3N2_Pandemic,aes(x=UNRATE,y=DATE),binwidth=20) + geom_bin2d()
```



This graph represents the unemployment rate during the H3N2 Pandemic and as shown it was much less harmful than covid has been. It also seems that the unemployment rate was consistently at about 3.6% which is a normal average of unemployment compared to every year up until covid, it seems that the covid-19 pandemic was something that has never been seen before in human history, due to a lack of precautions, covid-19 was able to spread around the entire globe within months of its arrival.

Filter dataset to display only data containing "walking" and "United States":

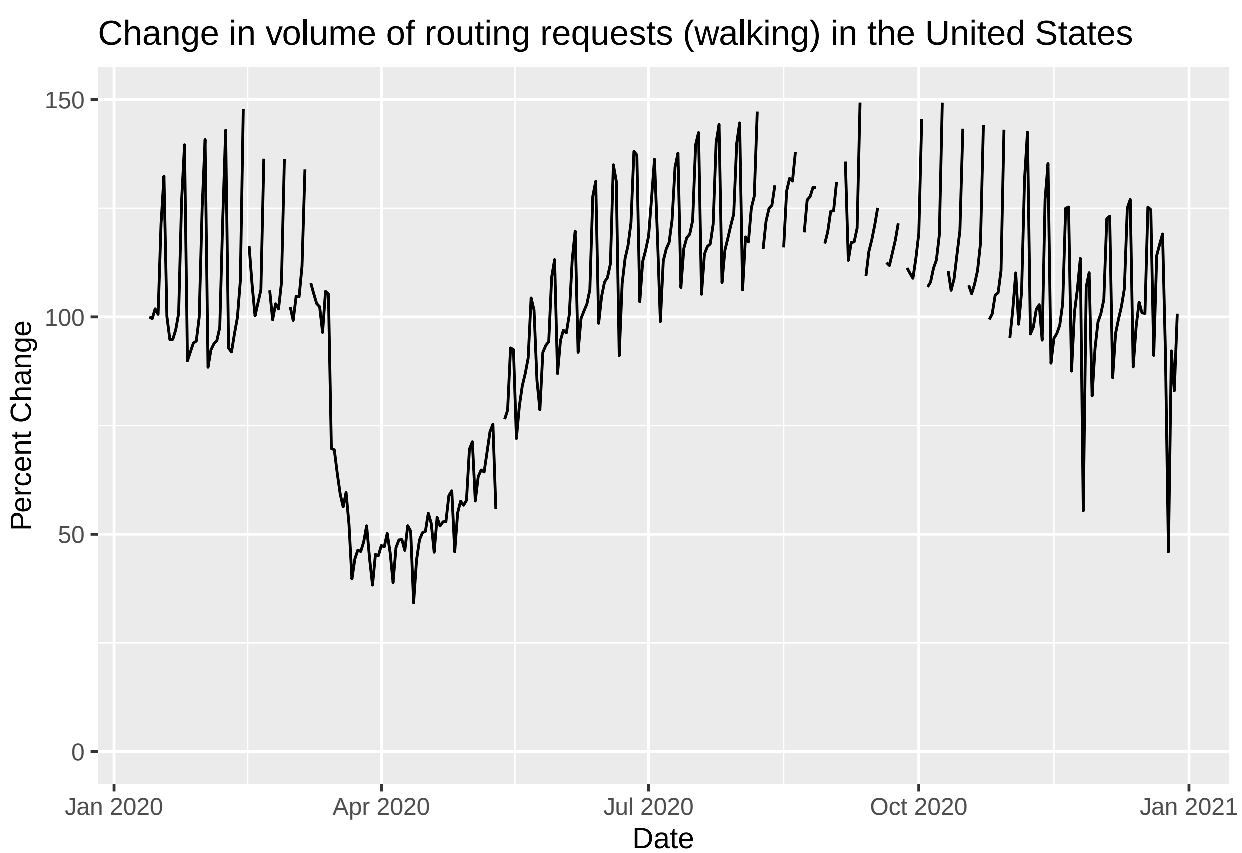
```
walking <- filter(apple_mobility2, region == "United States", transportation_type == "walking")
walking <- select(walking, -c(country, geo_type, region, transportation_type))
walking <- t(walking)
walking <- as.data.frame(walking)
head(walking,10)
```

```
##              V1
## X1.13.20 100.00
## X1.14.20  99.59
## X1.15.20 101.87
## X1.16.20 100.60
## X1.17.20 121.36
## X1.18.20 132.38
## X1.19.20 100.10
## X1.20.20  94.79
## X1.21.20  94.84
## X1.22.20  97.04
```

```r
walking      <- as.data.table(t(walking))
walking$date <- rownames(walking)
walking      <- melt(walking, id.vars=c("date"))
colnames(walking)[3] <- "percent"
walking$date <- as.numeric(gsub('X', '', walking$date, ignore.case=TRUE))
walking %<>%
  mutate(walking, variable=as.Date(variable, format="X%m.%d.%y"))
head(walking,10)
```

```
##     date   variable percent
##  1:    1 2020-01-13  100.00
##  2:    1 2020-01-14   99.59
##  3:    1 2020-01-15  101.87
##  4:    1 2020-01-16  100.60
##  5:    1 2020-01-17  121.36
##  6:    1 2020-01-18  132.38
##  7:    1 2020-01-19  100.10
##  8:    1 2020-01-20   94.79
##  9:    1 2020-01-21   94.84
## 10:    1 2020-01-22   97.04
```

```r
ggplot(walking, aes(x=variable, y=percent)) + geom_path() + ylim(0, 150) +
  ggtitle("Change in volume of routing requests (walking) in the United States") + xlab("Date") + ylab(
```



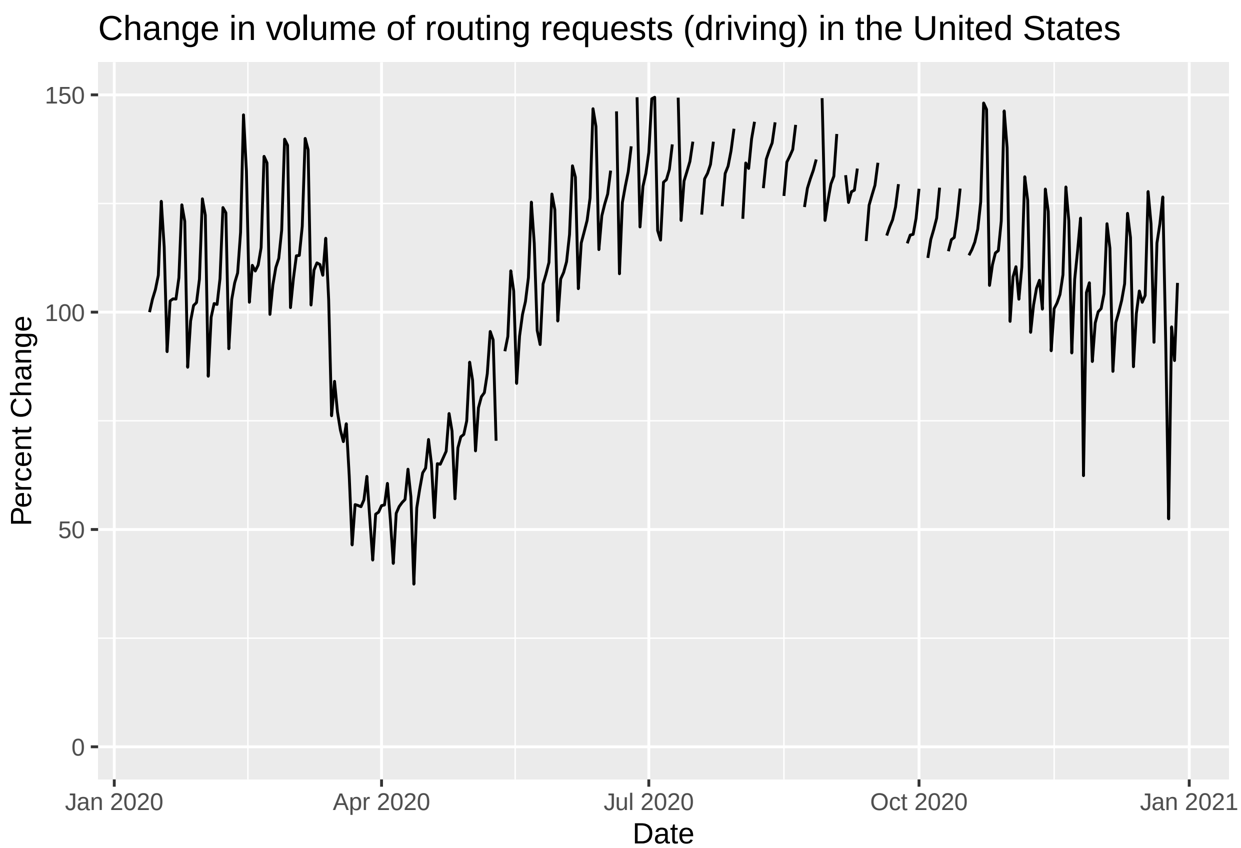
We can clearly see that the volume of routing requests for walking too a major dip when the pandemic happened and the stay at home order was issued. During this time obesity has increase in 12 states by about

%35 and this most likely happened when lockdown first started which is why the volume of routing requests fell by about 100%. It seems that covid-19 has made people gain significant weight because there was no proper procedures for going into public places. Many gyms closed and in result the obseity rates went up.

```
driving <- filter(apple_mobility2, region == "United States", transportation_type == "driving")
driving <- select(driving, -c(country, geo_type, region, transportation_type))
driving <- t(driving)
driving <- as.data.frame(driving)
driving <- as.data.table(t(driving))
driving$date <- rownames(driving)
driving      <- melt(driving, id.vars=c("date"))
colnames(driving)[3] <- "percent"
driving$date <- as.numeric(gsub('X', '', driving$date, ignore.case=TRUE))
driving %<>%
  mutate(driving, variable=as.Date(variable, format="X%m.%d.%y"))
driving
```

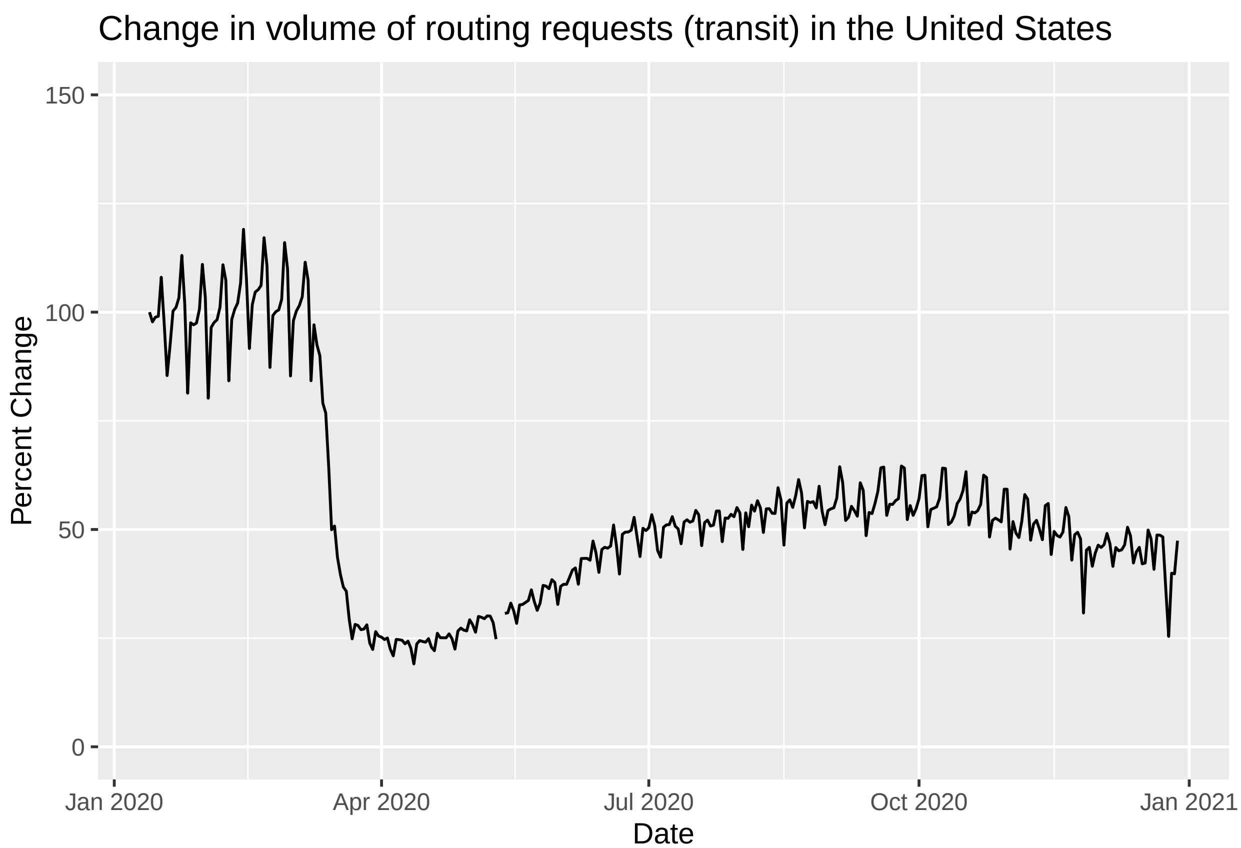
```
##       date   variable percent
##    1:    1 2020-01-13  100.00
##    2:    1 2020-01-14  102.97
##    3:    1 2020-01-15  105.19
##    4:    1 2020-01-16  108.48
##    5:    1 2020-01-17  125.51
##   ---
## 347:    1 2020-12-24   92.91
## 348:    1 2020-12-25   52.44
## 349:    1 2020-12-26   96.62
## 350:    1 2020-12-27   88.86
## 351:    1 2020-12-28  106.73
```

```
ggplot(driving, aes(x=variable, y=percent)) + geom_path() + ylim(0, 150) +
  ggtitle("Change in volume of routing requests (driving) in the United States") + xlab("Date") + ylab(
```



As seen in this graph, the requests for transportation dropped significantly during the periods between march - may. This is in direct correlation with when COVID became declared a global pandemic and places had started to go into lockdown. People stopped using public transportation as there was a high risk of catching the virus in places of high contact. Driving services such as Uber had suffered through the pandemic, losing around $1.1 billion over the last three months, with its adjusted net revenues down 20% compared to the third quarter of 2019

```
transit <- filter(apple_mobility2, region == "United States", transportation_type == "transit")
transit <- select(transit, -c(country, geo_type, region, transportation_type))
transit <- t(transit)
transit <- as.data.frame(transit)
transit      <- as.data.table(t(transit))
transit$date <- rownames(transit)
transit      <- melt(transit, id.vars=c("date"))
colnames(transit)[3] <- "percent"
transit$date <- as.numeric(gsub('X', '', transit$date, ignore.case=TRUE))
transit %<>%
  mutate(transit, variable=as.Date(variable, format="X%m.%d.%y"))
ggplot(transit, aes(x=variable, y=percent)) + geom_path() + ylim(0, 150) +
  ggtitle("Change in volume of routing requests (transit) in the United States") + xlab("Date") + ylab(
```



Based on the graphical results, we can see that before the pandemic there was a steady percent change in the volume of routing requests, hitting its peak around 125% and its low at around 75%. However, we see a significant decrease in percentage around Mid March, when the initial lock down had occurred. America was forced to socially distance, which meant to stay inside and reduce using public transportation, which explains the decrease in percentage during that time. The decrease in volume halts around late March to early April, where we see a small change in volume. Overtime, the volume starts to slowly increase, and is now holding a steady rate. There are many significant societal impacts that have occured due to this significant decrease in transit availability. Low income residents who rely on these transportations may have no choice but to walk or bike which will be difficult due to curfews and outdoor time limits. Furthermore, the vast majority of job opportunities within one hour of travel, is up to five times higher for those who own cars, compared to those who do not. This indefinitely denies low-income households from future job opportunities in their area, but will also increase unemployment rates as well as negatively impact long-term economic outcomes for both the city and the individuals.

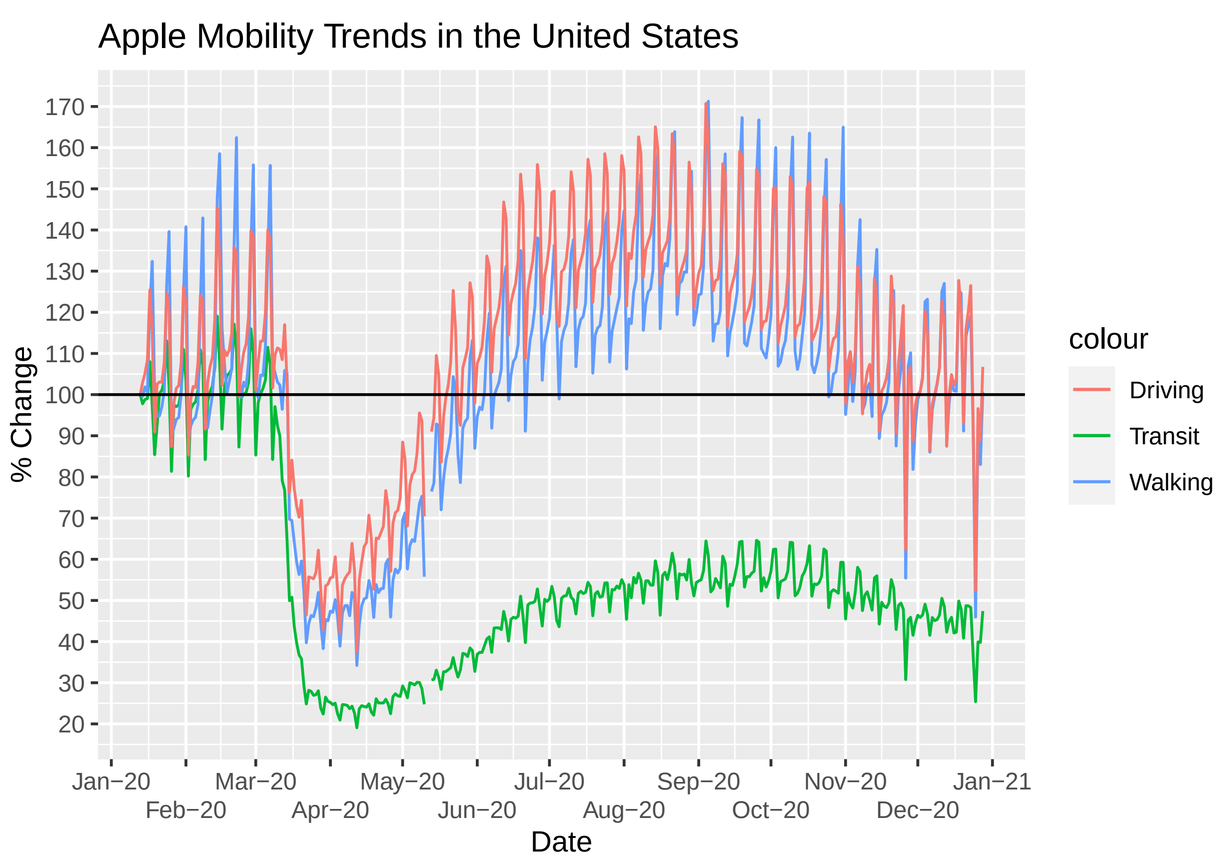
```
transit <- filter(apple_mobility2, region == "United States", transportation_type == "transit")
transit <- select(transit, -c(country, geo_type, region, transportation_type))
transit <- t(transit)
transit <- as.data.frame(transit)
transit      <- as.data.table(t(transit))
transit$date <- rownames(transit)
transit      <- melt(transit, id.vars=c("date"))
colnames(transit)[3] <- "percent"
transit$date <- as.numeric(gsub('X', '', transit$date, ignore.case=TRUE))
transit %<>%
  mutate(transit, variable=as.Date(variable, format="X%m.%d.%y"))
head(transit,10)
```

```
##      date   variable percent
##   1:     1 2020-01-13  100.00
##   2:     1 2020-01-14   97.77
##   3:     1 2020-01-15   98.84
##   4:     1 2020-01-16   99.04
##   5:     1 2020-01-17  108.03
##   6:     1 2020-01-18   97.41
##   7:     1 2020-01-19   85.40
##   8:     1 2020-01-20   92.39
##   9:     1 2020-01-21  100.27
##  10:     1 2020-01-22  101.06
```

```r
plot <- ggplot() +
geom_line(data=transit, aes(x=variable, y=percent, color='Transit')) +
geom_line(data=walking, aes(x=variable, y=percent, color='Walking')) +
geom_line(data=driving, aes(x=variable, y=percent, color='Driving')) + ylim(0, 150)
plot <- plot + ggtitle("Apple Mobility Trends in the United States")
plot <- plot + xlab("Date") + ylab("% Change") + scale_y_continuous(breaks=seq(0,200,10)) +
  scale_x_date(date_breaks = "months" , date_labels = "%b-%y", guide = guide_axis(n.dodge = 2))
```

```
## Scale for 'y' is already present. Adding another scale for 'y', which will
## replace the existing scale.
```

```r
plot <- plot + geom_hline(yintercept=100)
plot
```



Observing the data, we are able to see that walking and driving are the most commonly used methods of transportation with a high of approximately 163% and 145% before the pandemic respectively. Furthermore, walking had a low of approximately 93% while driving had a low of approximately 85%, meaning that before the pandemic these two methods had similar change in volume percentage. Transportation, although high, had a high of approximately 120% and a low of 80% before the pandemic. After the occurrence of the pandemic during mid March, there is a significant drop in the percentage of all three methods of transportation. The one most negatively affected was public transportation, as individuals avoided it as much as possible to reduce their chances of being affected by COVID-19. Late March to early April, is where we see an increase in both driving and walking. This is most likely due to the avoidance of public transportation, as it still does not show a significant increase until July.

Due to COVID-19, these three methods of transportation had long-term impacts on society, the economy, and personal health. Driving services such as Uber had suffered through the pandemic, losing around $1.1 billion over the last three months, with its adjusted net revenues down 20% compared to the third quarter of 2019. Furthermore, residents who need transit services due to low income or no access to a vehicle have little to no choice but to walk or bike. Job opportunities to these residents are impacted , ultimately increasing unemployment rates as well as decreasing long-term economic outcomes. Personal health was negatively impacted as well as there was an increase in obesity in 12 states by about 35%.

```
makeDF <- function(target, df, transport) {

  index <- df$region %in% target
  bestWalkLine <- df[index, ]
  bestWalkLine <- filter(bestWalkLine, transportation_type == transport)
  bestWalkLine <- select(bestWalkLine, -c(country, geo_type, transportation_type))
  bestWalkLine
  bestWalkLine <- t(bestWalkLine)
  bestWalkLine <- as.data.frame(bestWalkLine)
  colnames(bestWalkLine) <- bestWalkLine[1,]
  bestWalkLine <- bestWalkLine[-1,]
  bestWalkLine
  bestWalkLine$date <- rownames(bestWalkLine)
  bestWalkLine %<>%
    mutate(bestWalkLine, date=as.Date(date, format="X%m.%d.%y"))

  bestWalkLine[,1:(ncol(bestWalkLine)-1)] <- sapply(bestWalkLine[,1:(ncol(bestWalkLine)-1)],as.numeric)
  names(bestWalkLine)<-str_replace_all(names(bestWalkLine), c(" " = "_" , "," = "" ))
  bestWalkLine

  return(bestWalkLine)
}
```

```
walkingMean <- filter(apple_mobility, geo_type == "country/region", transportation_type == "walking")
walkingMean <- select(walkingMean,-c(geo_type,transportation_type,alternative_name, sub.region, country
```

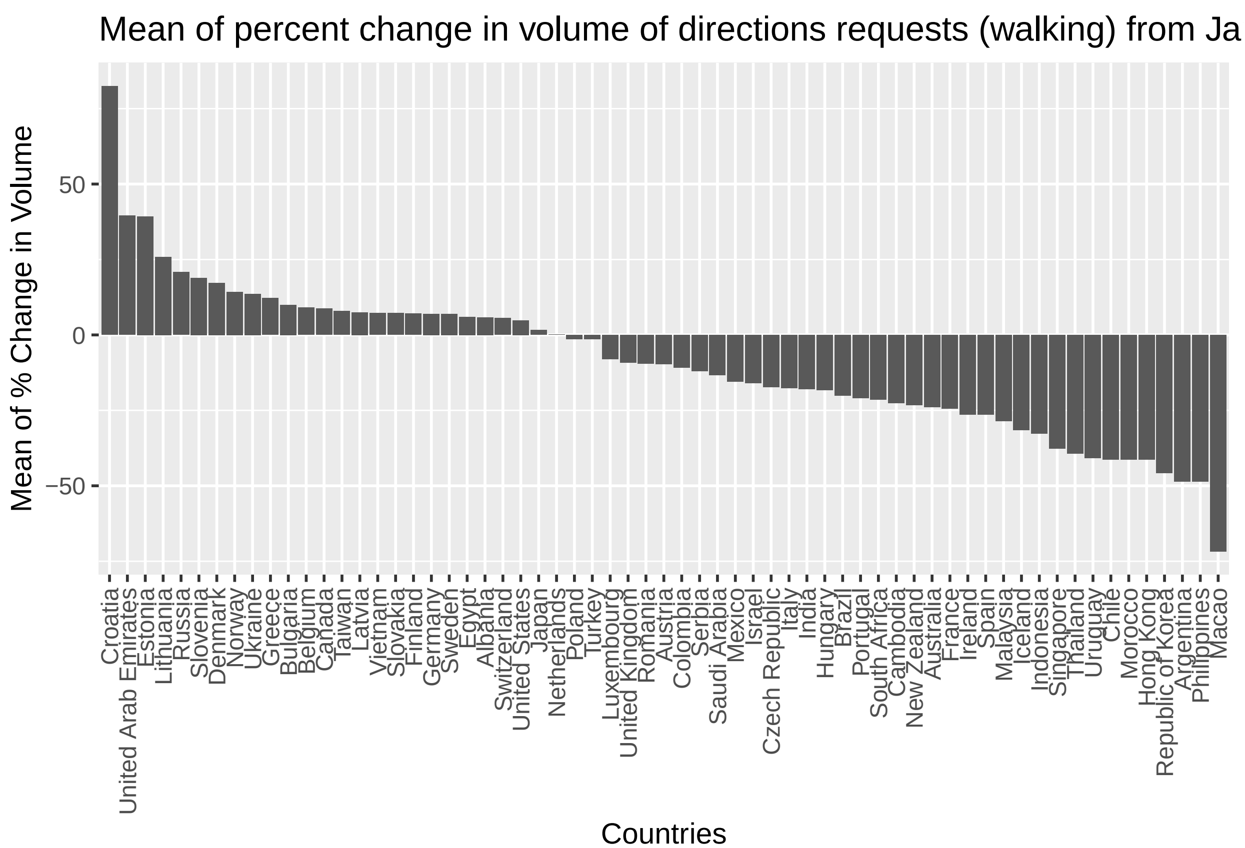
```
walkingMean$mean <- format(round(rowMeans(walkingMean[,-1], na.rm=TRUE) - 100, 2), nsmall = 2)
reqd <- as.vector(c("region","mean"))
walkingMean <- walkingMean[,reqd]
head(walkingMean,10)
```

```
##       region   mean
## 1    Albania   5.81
## 2  Argentina -48.55
## 3  Australia -23.88
```

```
## 4    Austria  -9.77
## 5    Belgium   9.05
## 6     Brazil -20.13
## 7   Bulgaria   9.98
## 8   Cambodia -22.63
## 9     Canada   8.76
## 10     Chile -41.29
```

```
walkingMean$mean <- as.numeric(as.character(walkingMean$mean))
```

```
p <-ggplot(walkingMean, aes(x=reorder(region,-mean), mean))+geom_bar(stat="identity", position=position_
p <- p + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  xlab("Countries") + ylab("Mean of % Change in Volume") + ggtitle("Mean of percent change in volume of
p
```



This graph is inconclusive.

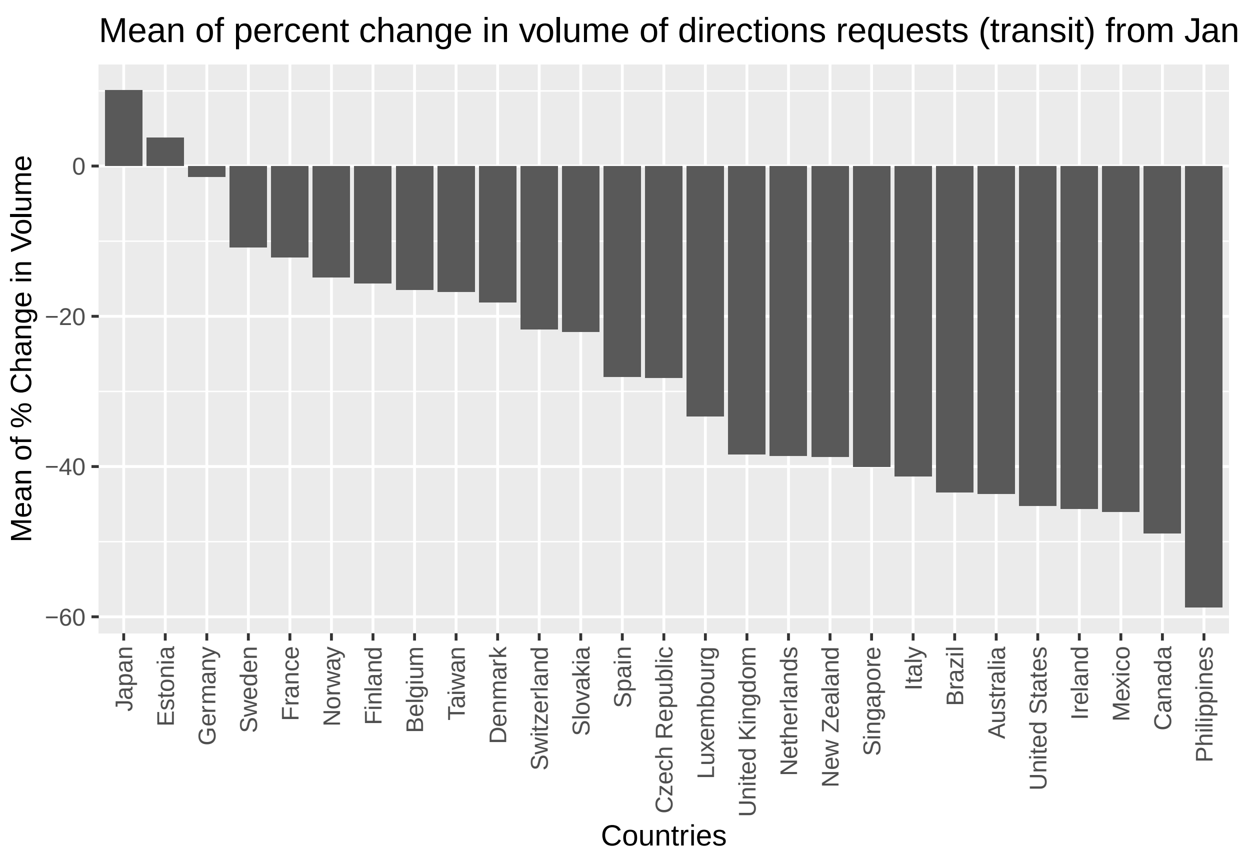
```
transitMean <- filter(apple_mobility, geo_type == "country/region", transportation_type == "transit")
transitMean <- select(transitMean,-c(geo_type,transportation_type,alternative_name, sub.region, country)


transitMean$mean <- format(round(rowMeans(transitMean[,-1], na.rm=TRUE) - 100, 2), nsmall = 2)
reqd <- as.vector(c("region","mean"))
transitMean <- transitMean[,reqd]
```

```
transitMean$mean <- as.numeric(as.character(transitMean$mean))
head(transitMean,10)
```

```
##            region   mean
## 1       Australia -43.60
## 2         Belgium -16.48
## 3          Brazil -43.42
## 4          Canada -48.90
## 5  Czech Republic -28.20
## 6         Denmark -18.15
## 7         Estonia   3.75
## 8         Finland -15.59
## 9          France -12.17
## 10        Germany  -1.40
```

```
p <-ggplot(transitMean, aes(x=reorder(region,-mean), mean))+geom_bar(stat="identity", position=position_
p <- p + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  xlab("Countries") + ylab("Mean of % Change in Volume") + ggtitle("Mean of percent change in volume of
p
```



As one can see from this graph, some countries had a positive percent change while others had a negative. Generally, a positive percent change is referencing an increase from its original value, while negative means a decrease. The majority of first world countries saw a positive change, an increase from their original value. Although the majority, some first world countries such as NZ and the United Kingdom saw a decrease. Moreover, countries from developing or underdeveloped regions saw a decrease in percent change, with

Macao seeing the heaviest decrease. People in the first world were driving more often instead of taking public transport to limit human contact and their chances of contracting the virus. Food services such as DoorDash and Ubereats became a lot more prominent as citizens chose to stay at home and limit contact with others. Places such as India where covid rates were among the highest, the use of transportation and food delivery was limited due to strict lockdown restrictions and even curfews. Causing a halt on economic growth and wealth accumulation of the citizens.

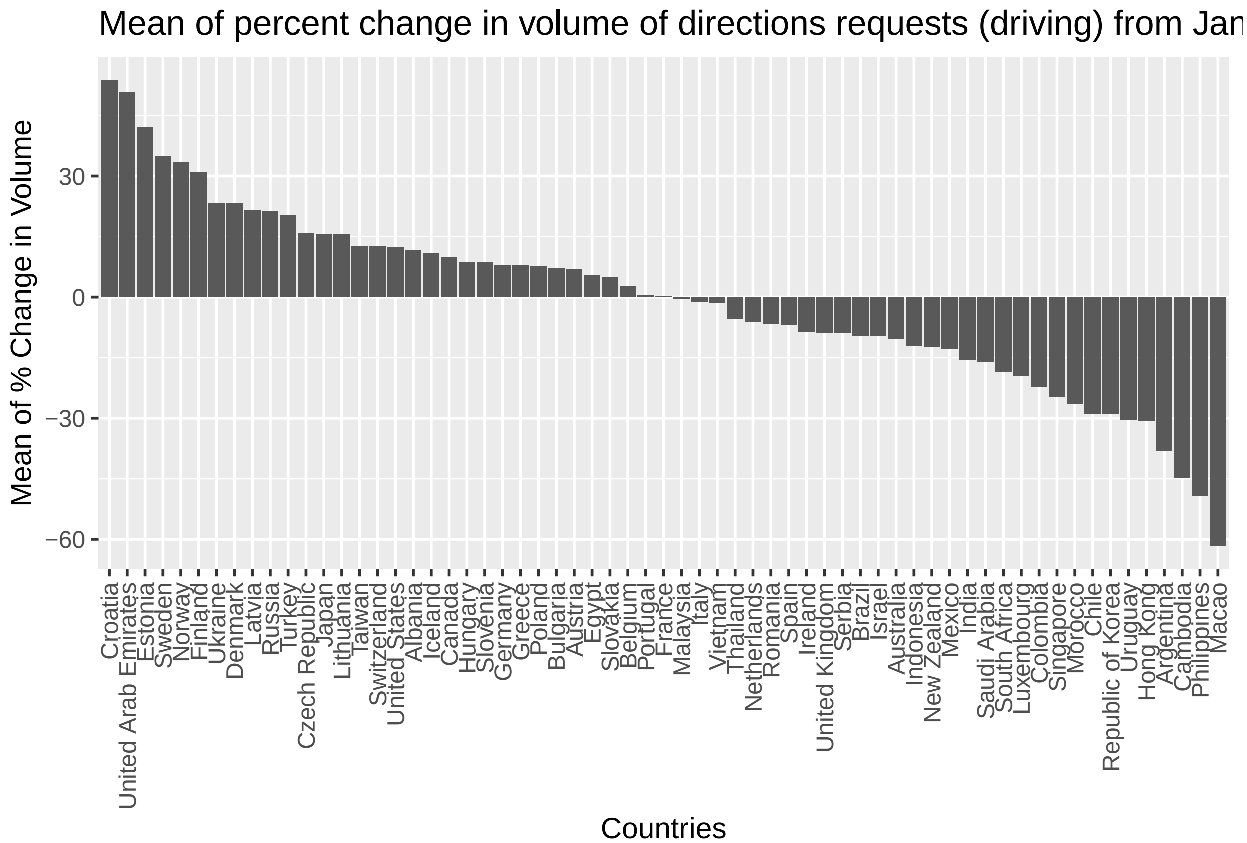
```
drivingMean <- filter(apple_mobility, geo_type == "country/region", transportation_type == "driving")
drivingMean <- select(drivingMean,-c(geo_type,transportation_type,alternative_name, sub.region, country)


drivingMean$mean <- format(round(rowMeans(drivingMean[,-1], na.rm=TRUE) - 100, 2), nsmall = 2)
reqd <- as.vector(c("region","mean"))
drivingMean <- drivingMean[,reqd]

drivingMean$mean <- as.numeric(as.character(drivingMean$mean))
head(drivingMean,10)
```

```
##       region   mean
## 1    Albania  11.60
## 2  Argentina -38.08
## 3  Australia -10.44
## 4    Austria   7.03
## 5    Belgium   2.79
## 6     Brazil  -9.53
## 7   Bulgaria   7.23
## 8   Cambodia -44.85
## 9     Canada   9.98
## 10     Chile -29.02
```

```
p <-ggplot(drivingMean, aes(x=reorder(region,-mean), mean))+geom_bar(stat="identity", position=position_
p <- p + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
 xlab("Countries") + ylab("Mean of % Change in Volume") + ggtitle("Mean of percent change in volume of d
p
```



This graph is inconclusive.

```
total <- merge(walkingMean, transitMean, by="region", all=TRUE)
total <- merge(total, drivingMean, by="region", all=TRUE)
colnames(total)[2:4] <- c("walking", "transit", "driving")
```

```
df.long<-melt(setDT(total))
```

```
## Warning in melt.data.table(setDT(total)): id.vars and measure.vars are
## internally guessed when both are 'NULL'. All non-numeric/integer/logical
## type columns are considered id.vars, which in this case are columns [region].
## Consider providing at least one of 'id' or 'measure' vars in future.
```

```
df.long$region <- factor(df.long$region, levels=unique(df.long$region))
head(df.long,10)
```

```
##         region variable  value
##   1:   Albania  walking   5.81
##   2: Argentina  walking -48.55
##   3: Australia  walking -23.88
##   4:   Austria  walking  -9.77
##   5:   Belgium  walking   9.05
##   6:    Brazil  walking -20.13
##   7:  Bulgaria  walking   9.98
##   8:  Cambodia  walking -22.63
```

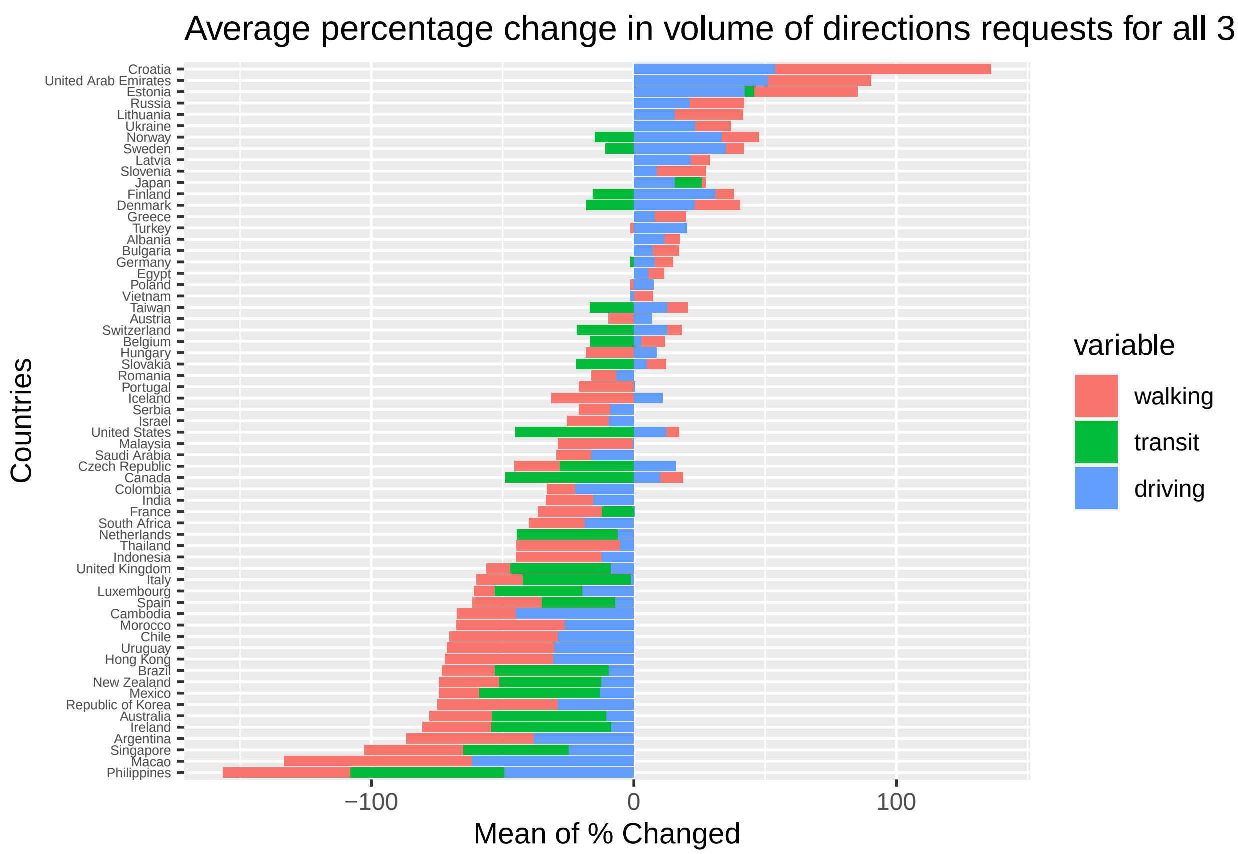
```
## 9:    Canada  walking   8.76
## 10:     Chile  walking -41.29
```

```
barPlot <- ggplot(df.long,aes(x=reorder(region, value, sum, na.rm=TRUE), y=value,fill=variable)) + geom_
 theme(axis.text.y = element_text(size=5, angle = 0))+
  xlab("Countries") + ylab("Mean of % Changed") + coord_flip() + ggtitle("Average percentage change in v
```

```
## Warning: Ignoring unknown parameters: stat
```

```
barPlot
```

```
## Warning: Removed 36 rows containing missing values (position_stack).
```



This graph is inconclusive as it does not provide any useful information.

Therefore, Due to COVID-19, Driving services such as Uber had suffered through the pandemic, losing around $1.1 billion over the last three months, with its adjusted net revenues down 20% compared to the third quarter of 2019. Furthermore, residents who need transit services due to low income or no access to a vehicle have little to no choice but to walk or bike. Job opportunities to these residents are impacted , ultimately increasing unemployment rates as well as decreasing long-term economic outcomes. Personal health was negatively impacted as well as there was an increase in obesity in 12 states by about 35%.