

Project 1 Analysis

Taylor Blair & Hans Niebuhr

Math 141, Week 5

Research Question

Is there a correlation between Covid precautions and number of cases?

Importing and Preparing Data

```
library(tidyverse)
library(ggplot2)
```

Loading data

The US Covid 19 behavior dataset comes from The Imperial College of London: <https://github.com/YouGov-Data/covid-19>

Its mission statement is the following:

YouGov has partnered with the Institute of Global Health Innovation (IGHI) at Imperial College London to gather global insights on people's behaviours in response to COVID-19. The research will cover 29 countries, interviewing around 21,000 people each week

The United States Covid dataset comes from The New York Times: <https://github.com/nytimes/covid-19-data>

Description from The NYTs:

The New York Times is releasing a series of data files with cumulative counts of coronavirus cases in the United States, at the state and county level, over time. We are compiling this time series data from state and local governments and health departments in an attempt to provide a complete record of the ongoing outbreak.

State population data comes from the Census: <https://www.census.gov/data>

No description provided

```
behavior_covid <- read_csv(file = "/home/courses/math141f20/Data/covid-19-behaviors/us_covid.csv")

state_covid_stats <- read_csv("https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-states")

state_pop_stats <- read_csv("nst-est2019-alldata.csv")
```

Preparing the State Cases Covid Data

The State by State Covid dataset is combined with the population dataset

```

state_pop_stats <- state_pop_stats %>%
  select(STATE, POPESTIMATE2019) %>%
  rename(fips=STATE)%>%
  rename(Population_2019=POPESTIMATE2019)

state_covid_stats <- state_covid_stats%>% #filter table to time range of behavior dataset
  filter(date>="2020-02-04") %>%
  filter(date<="2020-08-23") %>%
  merge(state_pop_stats, by = "fips") %>%
  subset(select=-(fips))

state_covid_stats <- state_covid_stats %>%
  mutate(
    per_cap_cases = round(cases/Population_2019*100000, 2),
    per_cap_deaths = round(deaths/Population_2019*100000, 2),
  ) %>%
  subset(select =-(Population_2019))

```

- **date**
– Date of the entry
- **state**
– State.
- **cases**
– “The total number of cases of Covid-19, including both confirmed and probable.”
- **deaths**
– “The total number of deaths from Covid-19, including both confirmed and probable.”
- **per_cap_cases**
– The total number of confirmed and probable cases divided by the estimate of the states 2019 population per 100,000.
- **per_cap_deaths**
– The total number of confirmed and probable deaths divided by the estimate of the states 2019 population per 100,000.

Covid Behavior Dataset

Clean Covid behavior dataset time frames

```

behavior_covid$endtime <- as.Date(behavior_covid$endtime, format = "%d/%m/%Y %H:%M")

behavior_covid = behavior_covid %>%
  rename(date=endtime)%>%
  subset(select =-(country))

```

Data Manipulation

Evaluating individual compliance and risk mitigation

The Covid Behavior dataset contains several variables that relate to safety precautions that individual takes.

```

behavior_covid <- behavior_covid %>%
  select(date, state, i12_health_1, i12_health_2,
         i12_health_3, i12_health_4, i12_health_5, i12_health_6,

```

```
i12_health_7, i12_health_8, i12_health_11, i12_health_12,
i12_health_13, i12_health_14, i12_health_15)
```

- **Date**
 - Date of an individuals survey %Y-%m-%d
- **State**
 - State survey conducted in.
- **i12_health**
 - 13 columns about Covid Precautions taken, from wearing a mask to outdoor interactions.
 - Categories are: *“Always, frequently, sometimes, rarely, not at all”*

Converting qualitative data to quantitative

It is not possible to take the average of labels, so the following have been converted.

```
#behavior_covid[8:20] <- lapply(behavior_covid[8:20], FUN)

behavior_covid[3:15][behavior_covid[3:15]=="Always"] <- "4"
behavior_covid[3:15][behavior_covid[3:15]=="Frequently"] <- "3"
behavior_covid[3:15][behavior_covid[3:15]=="Sometimes"] <- "2"
behavior_covid[3:15][behavior_covid[3:15]=="Rarely"] <- "1"
behavior_covid[3:15][behavior_covid[3:15]=="Not at all"] <- "0"
behavior_covid[3:15] <- sapply(behavior_covid[3:15], as.numeric)
```

Average of i12_health Questions

```
behavior_covid$avg_i12<- rowSums(behavior_covid[3:15])/12
```

Grouping by State and Date

```
behavior_covid <- behavior_covid %>%
  group_by(date, state) %>%
  summarise(mean_i_12=mean(avg_i12))
```

Aggregating with Covid Cases

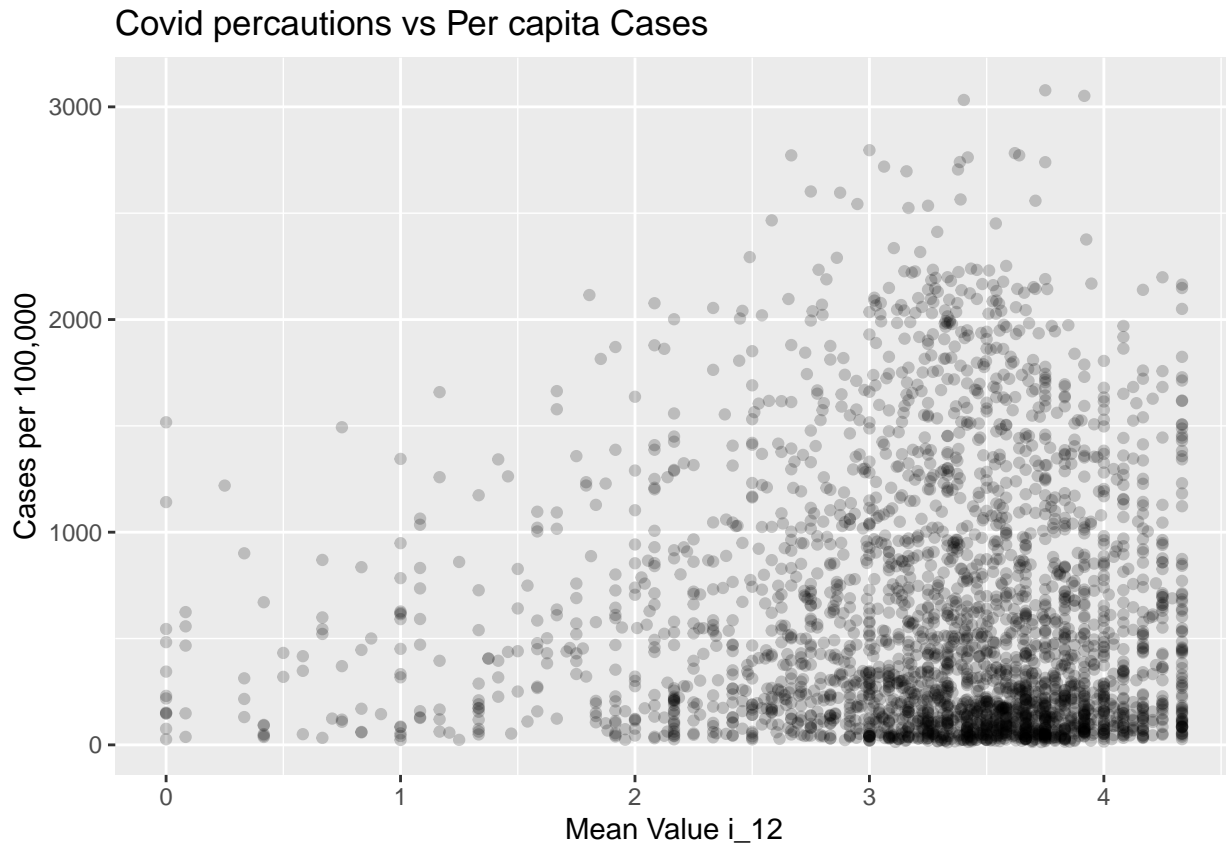
```
behavior_covid <- behavior_covid %>%
  merge(state_covid_stats,
        by=c("date", "state"))
```

Visualizing data

No filter scatter plot

A visualization with only per capita cases vs responses to i_12

```
ggplot(data=behavior_covid, mapping = aes(mean_i_12, per_cap_cases)) +
  geom_point(alpha=0.2) +
  labs(title = "Covid percautions vs Per capita Cases", x="Mean Value i_12", y="Cases per 100,000")
```



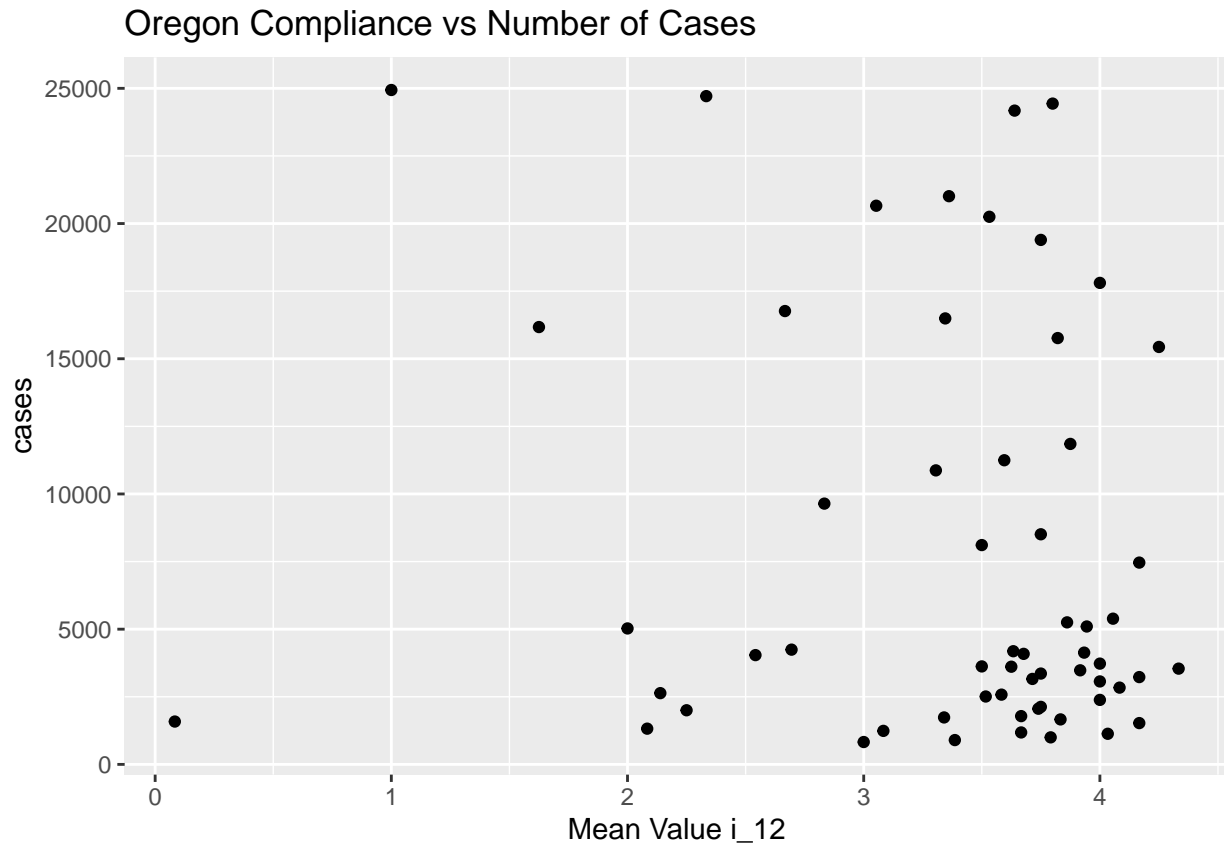
The graph appears chaotic with data points in the 4th quadrant. This could be because the data contains all dates and states are within this. The outlying data points do show an upward trend. This could be because as time increased so did individuals concerns

Single state

Because it is only a single state, and population estimates can be innacurate, we change to using the number of cases as opposed to the per capita rates.

```
Oregon <- behavior_covid%>% #filter table to time range of behavior dataset
  filter(state=="Oregon")

ggplot(data=Oregon,
  mapping = aes(mean_i_12, cases)) + geom_point(a=0.2) +
  labs(title = "Oregon Compliance vs Number of Cases", x="Mean Value i_12")
```



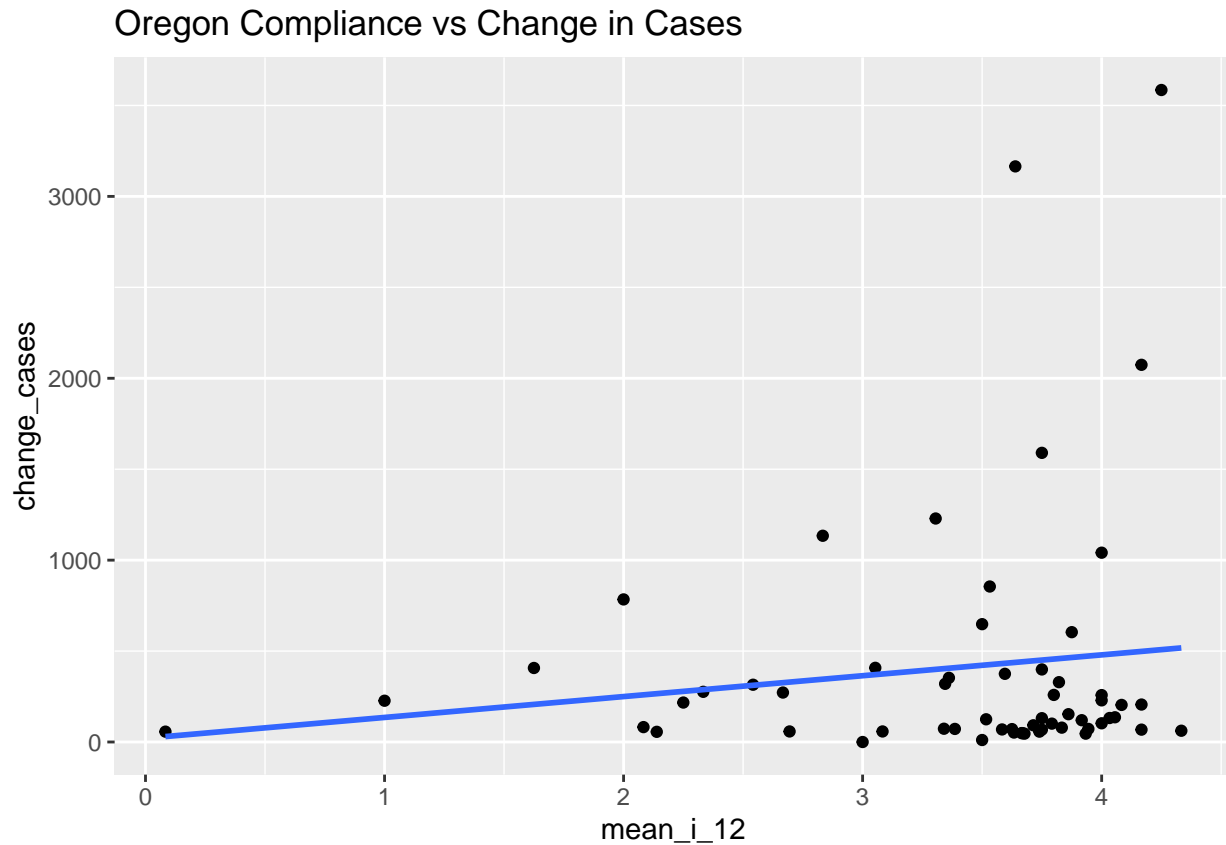
Because cases increase over time, it is possible to conclude that as the pandemic continues Oregonians have become less willing to follow the guidelines as demonstrated by the negative regression.

However, there has also been a slow in change of cases over time, which may have led Oregonians to feel as if the safety precautions are no longer necessary. So a look into the change in cases is necessary.

Change in cases over time

```
Oregon$change_cases<- c( 0, diff(Oregon$cases)) # calculates change in cases

ggplot(data=Oregon,
  mapping = aes(mean_i_12, change_cases)) + geom_point(a=0.2)+
  labs(title = "Oregon Compliance vs Change in Cases") + geom_smooth(method = lm, se = FALSE)
```



What this graph demonstrates is that although oregonians are centered around 3.5. Outliers tend to occur when there is a drastic change in cases, with the largest change in cases having one of the greatest `mean_i_12`.

Conclusion

There appears to be a coorelation between measures to protect ones self from becoming infected with covid and a change in cases. This makes sense considering as cases spike, so do individuals concerns about becoming infected leading individuals to display caution.