State Unemployment vs COVID

Robi Rahman

2021-04-19

Unemployment Data by State, January 2020 - March 2021

Importing and cleaning data on state unemployment rates during the coronavirus pandemic:

```
bls_unemployment_data <- read.csv("rawdata/Unemployment by state 2020-2021.csv")[,c(1,2,3,4,11)]

# List of state names:
state_names <- bls_unemployment_data$State[1:53]

# Initialize empty matrix
unemployment_data <- data.frame(matrix(ncol=55,nrow=15))

# Label matrix with state names
colnames(unemployment_data) <- c("Year", "Month", state_names)

# Initialize dates
unemployment_data$Year <- c(rep(2020,12),rep(2021,3))
unemployment_data$Month <- c(1:12, 1:3)


# Fill in matrix with data from the BLS file
for (m in 1:15) {
  for (s in 1:53) {
    unemployment_data[m,s+2]<-bls_unemployment_data[s+53*(m-1),5]
  }
}

# Save the cleaned data
write.csv(unemployment_data,"output/unemployment data.csv")
```

COVID Cases by State

Importing and cleaning data on state COVID case and death counts:

```
# Define helper functions to extract the year, month, and day of a COVID report:
year_num <- function(x) {
  as.numeric(substr(x,1,4))
}
month_num <- function(x) {
  as.numeric(substr(x,6,7))
}
day_num <- function(x) {
  as.numeric(substr(x,9,10))
}
```

```
# Read dataset of daily COVID cases and deaths for each state
covid_by_state <- read.csv("rawdata/COVID cases by state.csv")[,c(1,2,3,6,8,11)]

# Parse date strings from the file
covid_by_state$submission_date <- as.Date(covid_by_state$submission_date, format = "%m/%d/%Y")

# Get the year, month, and day using the helper functions:
covid_by_state$year <- year_num(covid_by_state$submission_date)
covid_by_state$month <- month_num(covid_by_state$submission_date)
covid_by_state$day <- day_num(covid_by_state$submission_date)

# Sort dataframe by location and date
attach(covid_by_state)
covid_by_state <- covid_by_state[order(year, month, day, state),]
detach(covid_by_state)
```

Now we need to group and sum the daily cases in each state by month:

```
covid_data <- covid_by_state %>%
  group_by(state, year, month) %>%
  summarize(cases=sum(new_case),deaths=sum(new_death))
```

```
## `summarise()` has grouped output by 'state', 'year'. You can override using the `.groups` argument.
```

We now have COVID data by state and month, but need to organize it.

```
# Copy the monthly unemployment data table:
cases_data <- unemployment_data[,1:2]
cases_data[16,] <- c(2021,4)
deaths_data <- unemployment_data[,1:2]
deaths_data[16,] <- c(2021,4)

# Fill in the monthly COVID case and deaths dataframes:
for (s in 1:60) {
  cases_data[s+2] <- rep(0,16)
  colnames(cases_data)[s+2] <- covid_data$state[1+16*(s-1)]
  deaths_data[s+2] <- rep(0,16)
  colnames(deaths_data)[s+2] <- covid_data$state[1+16*(s-1)]
  for (m in 1:16) {
    cases_data[m,s+2] <- covid_data[1+16*(s-1)+m,4]
    deaths_data[m,s+2] <- covid_data[1+16*(s-1)+m,5]
  }
}
```

Remove American Samoa, Federated States of Micronesia, Guam, MP, PW, RMI, Virgin Islands, and Los Angeles County. (I'm retaining New York City because it is present in both the unemployment and COVID cases/deaths datasets.)

```
cases_data <- cases_data[c(-16),c(-6,-14,-16,-31,-47,-48,-50,-57)]
deaths_data <- deaths_data[c(-16),c(-6,-14,-16,-31,-47,-48,-50,-57)]
unemployment_data <- unemployment_data[,c(-8)]
```

Sorting the unemployment data by abbreviation to match the case data:

```
colnames(unemployment_data) <- c(
  "Year","Month","AL","AK","AZ","AR","CA","CO","CT","DE","DC","FL","GA","HI",
  "ID","IL","IN","IA","KS","KY","LA","ME","MD","MA","MI","MN","MS","MO","MT",
```

```
  "NE","NV","NH","NJ","NM","NY","NYC","NC","ND","OH","OK","OR","PA","RI","SC",
  "SD","TN","TX","UT","VT","VA","WA","WV","WI","WY"
)
unemployment_data <- unemployment_data[,
                                       c(1,2,4,3,6,5,7:9,11,10,12:14,18,15:17,
                                         19:21,24,23,22,25:26,28,27,29,37:38,
                                         30,32,33:34,31,35,36,39:48,
                                         50,49,51,53,52,54)
                                       ]
```

We can subtract the previous month's unemployment rate from the next month's unemployment rate to find the monthly change in unemployment:

```
unemployment_changes <- unemployment_data
unemployment_changes[1:14,3:54] <- unemployment_changes[2:15,3:54] - unemployment_changes[1:14,3:54]
unemployment_changes <- unemployment_changes[c(-15),]
```

Now let's find the correlation between COVID cases and deaths, and state unemployment rates.

Warning: this code chunk displays 104 scatterplots!

```
# for (i in 1:52) {
#   cases <- cases_data[1:14,i+2]
#   deaths <- deaths_data[1:14,i+2]
#   unemp_chg <- unemployment_changes[,i+2]
#   cases_vs_unemp <- lm(unemp_chg ~ cases)
#   deaths_vs_unemp <- lm(unemp_chg ~ deaths)
#   plot(cases,unemp_chg, main=paste(colnames(unemployment_changes)[i+2],"unemployment change vs cases"
#   abline(cases_vs_unemp$coefficients[[1]],cases_vs_unemp$coefficients[[2]])
#   plot(deaths, unemp_chg, main=paste(colnames(unemployment_changes)[i+2],"unemployment change vs deat
#   abline(deaths_vs_unemp$coefficients[[1]],deaths_vs_unemp$coefficients[[2]])
# }
```