

4-exploration

Ashley Clarke

12/13/2021

Need to use the following R packages and functions:

```
# load libraries
library(kableExtra)           # for printing tables
library(cowplot)              # for side by side plots
library(lubridate)            # for dealing with dates
library(maps)                 # for creating maps
library(tidyverse)
library(ggcorrplot)           # for correlation plots

## Rows: 585 Columns: 217

## -- Column specification -----
## Delimiter: "\t"
## dbl  (216): AHECONS, AWHMAN, AWOTMAN, CES0600000006, CES1000000006, CES10000...
## date  (1): date

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## Rows: 147 Columns: 217

## -- Column specification -----
## Delimiter: "\t"
## dbl  (216): AHECONS, AWHMAN, AWOTMAN, CES0600000006, CES1000000006, CES10000...
## date  (1): date

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

test_obs = nrow(econ_train)
train_obs = nrow(econ_test)
total_obs = test_obs + train_obs
num_features = ncol(econ_train)
```

#The test data set has 585 observations, while the train data set has 147 observations. The data set represents 732 different months of data from June 1960 to December 2020. There are 217 features

Variation in response variable

```
histogram_unemploy = econ_train %>%
  ggplot(aes(x = UNRATE)) +
```

```

geom_histogram(bins = 15, fill = "grey", col = "black") +
labs(x = "Unemployment Rate (%)",
     y = "Count") +
# add vertical line at the median value for unemployment rate
geom_vline(xintercept = median(econ_train$UNRATE), color = "#f04546") +
theme_bw()

# save the plot
ggsave(filename = "~/Desktop/STAT471/unemployment-project/results/histogram_unemploy.png",
       plot = histogram_unemploy,
       device = "png",
       width = 8,
       height = 3)

# which 10 months have the highest unemployment rates
# do these correspond with recessions?
top_10_unemploy = econ_train %>%
  select(date, UNRATE) %>%
  arrange(desc(UNRATE)) %>%
  head(10)

top_10_unemploy %>%
  kable(format = "latex", row.names = NA,
        booktabs = TRUE, digits = 1) %>%
  kable_styling(position = "center") %>%
  save_kable(file =
    "~/Desktop/STAT471/unemployment-project/results/top_10_unemploy.pdf",
    self_contained = T)

```

Calculate mean unemployment

```

# All data
mean_all_years = econ_train %>%
  summarise("Time" = "1960-2020", "Unemployment Rate" = mean(UNRATE))

# After 2010
mean_last_10years = econ_train %>%
  filter(date > as.Date("2010-01-01")) %>%
  summarise("Time" = "2010-2020", "Unemployment Rate" = mean(UNRATE))

mean_unemploy = rbind(mean_all_years, mean_last_10years)

mean_unemploy %>%
  kable(format = "latex", row.names = NA,
        booktabs = TRUE, digits = 2) %>%
  kable_styling(position = "center") %>%
  save_kable(file =
    "~/Desktop/STAT471/unemployment-project/results/mean-unemployment-chart.pdf",
    self_contained = T)

#Plot unemployment by year

```

```

# set colors to use for graphs
cols <- c("Total U.S." = "#3591d1", "Mean Unemployment (1972-2021)" = "#f04546")

# plot both unemployment rates for all years
all_year_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x=date, y=UNRATE, colour = "Total U.S.)) +
  geom_hline(yintercept = median(econ_train$UNRATE),
             colour = "#f04546", linetype='dashed') +
  ggtitle("Total U.S. Unemployment (1960-2020)") +
  theme_bw() +
  scale_colour_manual(name="Metric", values=cols) +
  labs(
    x = "Date",
    y = "Unemployment Rate"
  )
# save the plot
ggsave(filename = "~/Desktop/STAT471/unemployment-project/results/all-year-comparison-plot.png",
        plot = all_year_plot,
        device = "png",
        width = 8,
        height = 3)

```

Consider Phillips Curve - plot Unemployment vs. Inflation

Does the Phillips Curve hold true for either party?

```

# Plot phillips curve by year
phillips_curve_US = econ_train %>%
  summarise(year = year(date), UNRATE, FPCPITOTLZGUSA) %>%
  group_by(year) %>%
  summarise(unemployment_rate = mean(UNRATE),
            inflation = mean(FPCPITOTLZGUSA)) %>% ungroup() %>%

  ggplot() +
  geom_line(mapping = aes(x=unemployment_rate, y=inflation), color = "#3591d1") +
  ggtitle("Inflation vs. Unemployment Rate") +
  theme_bw() +
  labs(
    x = "Unemployment Rate",
    y = "Inflation"
  )

# save the plot
ggsave(filename =
  "~/Desktop/STAT471/unemployment-project/results/phillips-curve-us-plot.png",
  plot = phillips_curve_US,
  device = "png",
  width = 8,
  height = 3)

```

Plot unemployment rate vs. federal funds rate

```
# plot both unemployment rates
fed_funds_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x= FEDFUNDS,y=UNRATE), color = "#3591d1") +
  ggtitle("Unemployment Rate vs. Federal Funds Rate") +
  theme_bw() +
  labs(
    x = "Federal Funds Rate",
    y = "Unemployment Rate"
  )
# save the plot
ggsave(filename =
  "~/Desktop/STAT471/unemployment-project/results/fed-funds-plot.png",
  plot = fed_funds_plot,
  device = "png",
  width = 8,
  height = 3)
```

Look at variables related gto manufacturing, mining/logging, construction, information

```
# examine unemployment by thousands of employees

# goods producing plot
employees_goods_producing_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x= CEU0600000001, y = UNRATE), color = "#3591d1") +
  ggtitle("Manufacturing") +
  theme_bw() +
  theme(axis.title.x = element_text(size = 8)) +
  theme(axis.title.y = element_text(size = 8)) +
  labs(
    x = "Goods Producing Employees (thousands)",
    y = "Unemployment Rate"
  )

# logging plot
employees_logging_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x= CES1011330001, y = UNRATE), color = "#3591d1") +
  ggtitle(" Mining/Logging") +
  theme_bw() +
  theme(axis.title.x = element_text(size = 8)) +
  theme(axis.title.y = element_text(size = 8)) +
  labs(
    x = "Logging Employees (thousands)",
    y = "Unemployment Rate"
  )

# shipping and boating
employees_shipping_boating_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x= CES3133660001, y = UNRATE), color = "#3591d1") +
```

```

ggtitle("Construction") +
theme_bw() +
theme(axis.title.x = element_text(size = 8)) +
theme(axis.title.y = element_text(size = 8)) +
labs(
  x = "Shipping and Boating Employees (thousands)",
  y = "Unemployment Rate"
)

# information plot
employees_information_plot = econ_train %>%
ggplot() +
geom_line(mapping = aes(x= CEU50000000001, y = UNRATE), color = "#3591d1") +
ggtitle("Information") +
theme_bw() +
theme(axis.title.x = element_text(size = 8)) +
theme(axis.title.y = element_text(size = 8)) +
labs(
  x = "Information Employees (thousands)",
  y = "Unemployment Rate"
)

# information plot after 2000, since likely meaningless before then
employees_information_plot_after_2000 = econ_train %>%
filter(date > as.Date("2000-01-01")) %>%
ggplot() +
geom_line(mapping = aes(x= CEU50000000001, y = UNRATE), color = "#3591d1") +
ggtitle("Information (2000-2020)") +
theme_bw() +
theme(axis.title.x = element_text(size = 8)) +
theme(axis.title.y = element_text(size = 8)) +
labs(
  x = "Information Employees (thousands)",
  y = "Unemployment Rate"
)

# federal plot, suggests numbrer might increase during high unemployment
employees_federal_plot = econ_train %>%
ggplot() +
geom_line(mapping = aes(x= CES9091000001, y = UNRATE), color = "#3591d1") +
ggtitle("Federal") +
theme_bw() +
theme(axis.title.x = element_text(size = 8)) +
theme(axis.title.y = element_text(size = 8)) +
labs(
  x = "Federal Employees (thousands)",
  y = "Unemployment Rate"
)

employees_plot = plot_grid(employees_goods_producing_plot , employees_logging_plot, employees_shipping_l

## Warning in as_grob.default(plot): Cannot convert object of class numeric into a
## grob.

```

```

# save the plot
ggsave(filename =
  "~/Desktop/STAT471/unemployment-project/results/employees_plot.png",
  plot = employees_plot,
  device = "png",
  width = 8,
  height = 6)

```

examine unemployment by hours of work, issue is still classified as employed even if hours change

```

# hours manufacturing plot
hours_production_manufacturing_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x= AWHMAN, y = UNRATE), color = "#3591d1") +
  ggtitle("Manufacturing") +
  theme_bw() +
  theme(axis.title.x = element_text(size = 8)) +
  theme(axis.title.y = element_text(size = 8)) +
  labs(
    x = "Hours Production",
    y = "Unemployment Rate"
  )

```

```

# hours mining and logging plot
hours_production_mining_logging_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x= CES1000000007, y = UNRATE), color = "#3591d1") +
  ggtitle("Mining/Logging") +
  theme_bw() +
  theme(axis.title.x = element_text(size = 8)) +
  theme(axis.title.y = element_text(size = 8)) +
  labs(
    x = "Hours Production",
    y = "Unemployment Rate"
  )

```

```

# hours construction plot
hours_production_construction_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x= CEU2000000007, y = UNRATE), color = "#3591d1") +
  ggtitle("Construction") +
  theme_bw() +
  theme(axis.title.x = element_text(size = 8)) +
  theme(axis.title.y = element_text(size = 8)) +
  labs(
    x = "Hours Production",
    y = "Unemployment Rate"
  )

```

```

hours_production_plot = plot_grid(hours_production_manufacturing_plot, hours_production_mining_logging_

```

```

# save the plot

```

```
ggsave(filename =
  "~/Desktop/STAT471/unemployment-project/results/hours_production_plot.png",
  plot = hours_production_plot,
  device = "png",
  width = 8,
  height = 6)
```

```
# women employees durable goods
```

```
women_employees_durable_goods_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x= CES3100000010, y = UNRATE), color = "#3591d1") +
  ggtitle("Female- Durable") +
  theme_bw() +
  theme(axis.title.x = element_text(size = 8)) +
  theme(axis.title.y = element_text(size = 8)) +
  labs(
    x = "Female Employees Durable Goods",
    y = "Unemployment Rate"
  )
```

```
# women employees durable goods
```

```
women_employees_durable_goods_plot_after_2000 = econ_train %>%
  filter(date > as.Date("2000-01-01")) %>%
  ggplot() +
  geom_line(mapping = aes(x= CES3100000010, y = UNRATE), color = "#3591d1") +
  ggtitle("Female- Durable (2000-2020)") +
  theme_bw() +
  theme(axis.title.x = element_text(size = 8)) +
  theme(axis.title.y = element_text(size = 8)) +
  labs(
    x = "Female Employees Durable Goods",
    y = "Unemployment Rate"
  )
```

```
# women employees non durable goods
```

```
women_employees_nondurable_goods_plot = econ_train %>%
  ggplot() +
  geom_line(mapping = aes(x= CES3200000010, y = UNRATE), color = "#3591d1") +
  ggtitle("Female Employees - Nondurable") +
  theme_bw() +
  theme(axis.title.x = element_text(size = 8)) +
  theme(axis.title.y = element_text(size = 8)) +
  labs(
    x = "Female- Nondurable Goods",
    y = "Unemployment Rate"
  )
```

```
# women employees non durable goods
```

```
women_employees_nondurable_goods_plot_after_2000 = econ_train %>%
  filter(date > as.Date("2000-01-01")) %>%
  ggplot() +
  geom_line(mapping = aes(x= CES3200000010, y = UNRATE), color = "#3591d1") +
  ggtitle("Female- Nondurable (2000-2020)") +
  theme_bw() +
  theme(axis.title.x = element_text(size = 8)) +
```

```

theme(axis.title.y = element_text(size = 8)) +
labs(
  x = "Female Employees Nondurable Goods",
  y = "Unemployment Rate"
)

female_plots = plot_grid(women_employees_durable_goods_plot, women_employees_durable_goods_plot_after_2008,
                        women_employees_nondurable_goods_plot, women_employees_nondurable_goods_plot_after_2008)

# save the plot
ggsave(filename =
  "~/Desktop/STAT471/unemployment-project/results/female_plots .png",
  plot = female_plots ,
  device = "png",
  width = 8,
  height = 6)

```