Communicating Change Assignment

Nick Sun

May 12, 2019

Looking at the data source, we see that `employed` includes total non-farm worker employees (roughly 80% of the workforce that contributes to the GDP of the United States). The units is in thousands of persons.

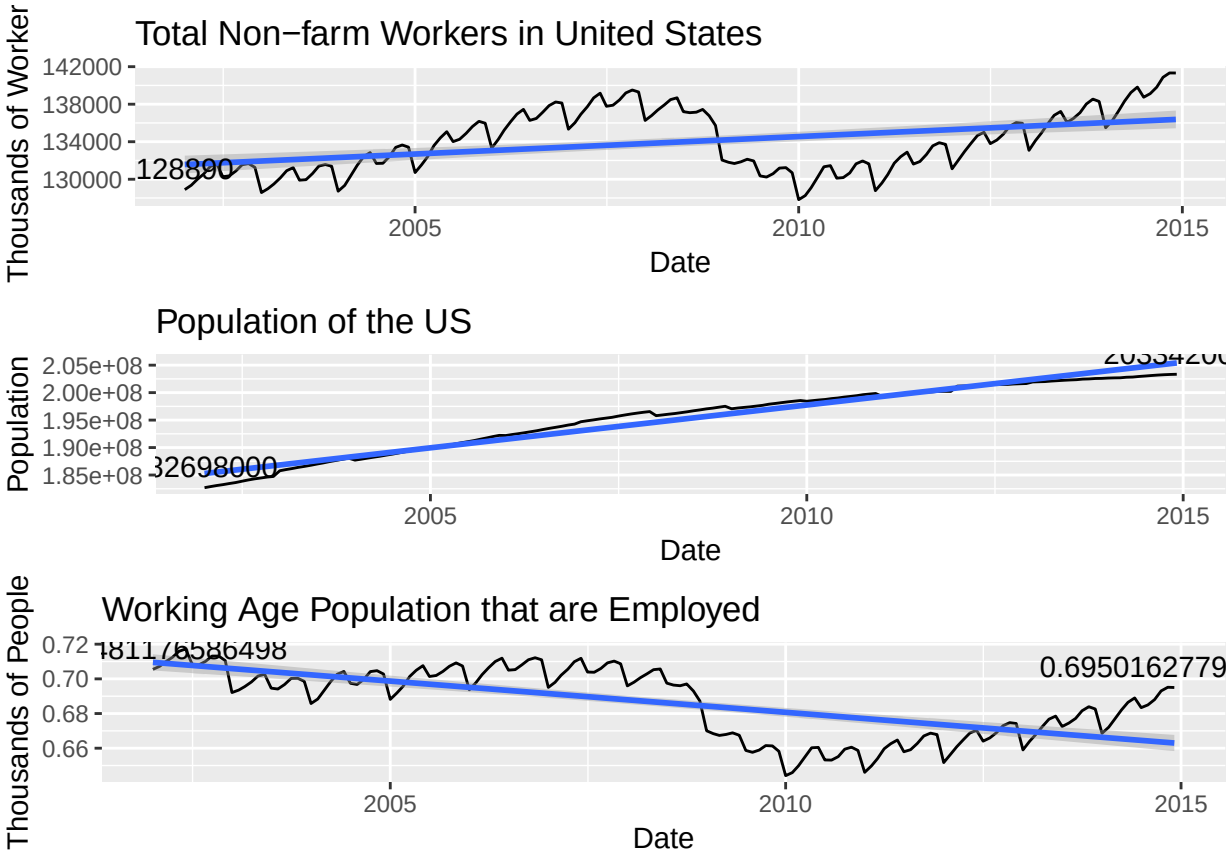
The data for the working age population of the country has persons as the unit. If we make a new column `employed_scaled` we will have to remember to correct for this difference in units.

Filtering on `Date` datatypes can be done with the `lubridate::as.Date` function.

```
pop_emp %>%
  filter(date >= as.Date("2002-01-01") & date <= as.Date("2014-12-31")) -> pop_emp_recent
```

```
firstnlast <- pop_emp_recent[c(1, length(pop_emp_recent$date)),]

# firstnlast %>%
#   mutate(employed_scaled = employed_scaled / 1000000) -> firstnlast
#
# pop_emp_recent %>%
#   mutate(employed = employed / 1000) -> pop_emp_recent

g1 <- ggplot() +
  geom_line(pop_emp_recent, mapping = aes(x = date, y = employed)) +
  geom_smooth(pop_emp_recent, mapping = aes(x = date, y = employed), method = "lm") +
  geom_text(firstnlast, mapping = aes(x = date,
                                      y = employed,
                                      label = employed),
            vjust = -.5) +
  labs(title = "Total Non-farm Workers in United States",
       x = "Date",
       y = "Thousands of Workers")

g2 <- ggplot() +
  geom_line(pop_emp_recent, mapping = aes(x = date, y = pop)) +
  geom_smooth(pop_emp_recent, mapping = aes(x = date, y = pop), method = "lm") +
  geom_text(firstnlast, mapping = aes(x = date,
                                      y = pop,
                                      label = pop),
            vjust = -.5) +
  labs(title = "Population of the US",
       x = "Date",
       y = "Population")

g3 <- ggplot() +
  geom_line(pop_emp_recent, mapping = aes(x = date, y = percent_emp)) +
  geom_smooth(pop_emp_recent, mapping = aes(x = date, y = percent_emp), method = "lm") +
  geom_text(firstnlast, mapping = aes(x = date,
                                      y = percent_emp,
                                      label = percent_emp),
            vjust = -.5) +
```

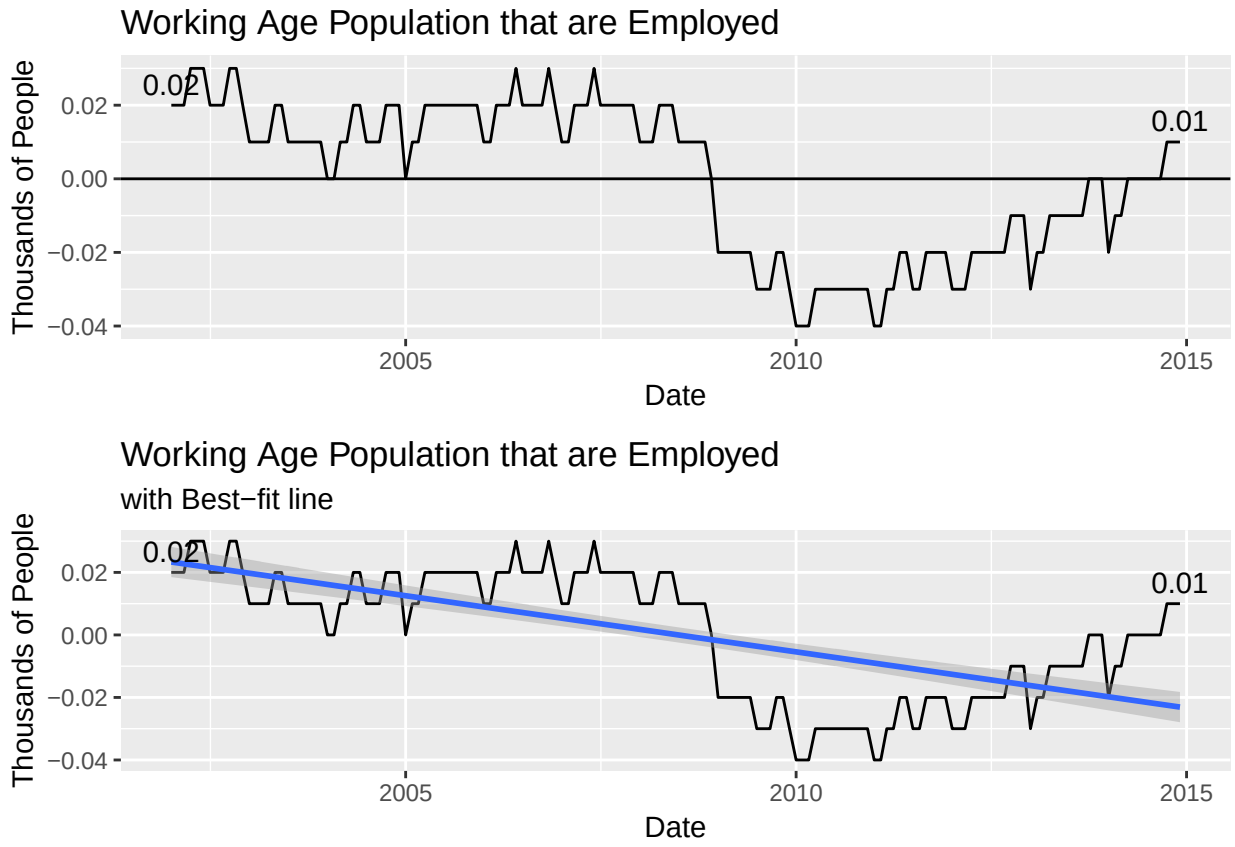
```
  labs(title = "Working Age Population that are Employed",
       x = "Date",
       y = "Thousands of People")

grid.arrange(g1, g2, g3, nrow = 3)
```



Now we have to find the average percent employed between January 2002 and December 2014 and then compare each percentage score with that average percentage.

```
average_percent_emp <- mean(pop_emp_recent$percent_emp)
pop_emp_recent %>%
  mutate(percent_change = round(percent_emp - average_percent_emp, 2)) -> pop_emp_recent
firstnlast %>%
  mutate(percent_change = round(percent_emp - average_percent_emp, 2)) -> firstnlast

g4 <- ggplot() +
  geom_line(pop_emp_recent, mapping = aes(x = date, y = percent_change)) +
  geom_hline(yintercept = 0) +
  geom_text(firstnlast, mapping = aes(x = date,
                                      y = percent_change,
                                      label = percent_change),
            vjust = -.5) +
  labs(title = "Working Age Population that are Employed",
       x = "Date",
       y = "Thousands of People")
```

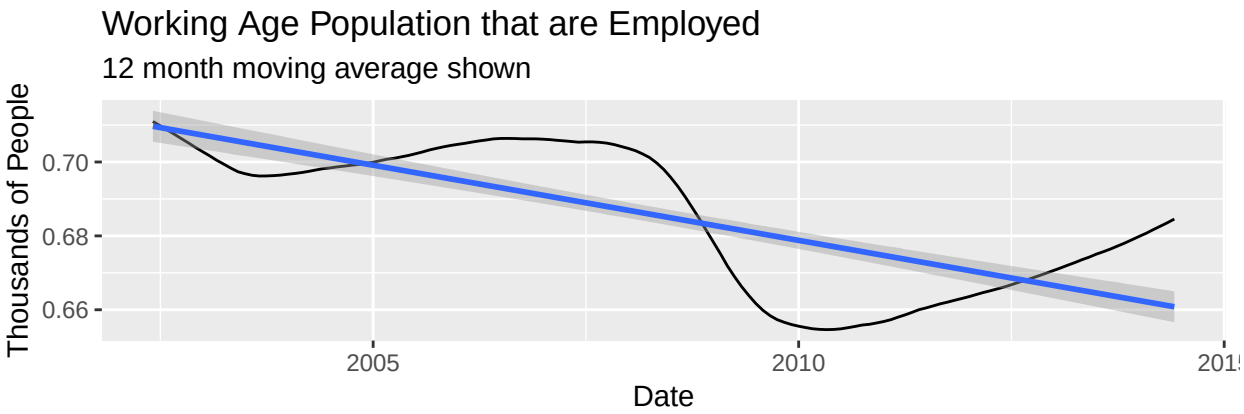
```
g5 <- ggplot() +
  geom_line(pop_emp_recent, mapping = aes(x = date, y = percent_change)) +
  geom_smooth(pop_emp_recent,
              mapping = aes(x = date, y = percent_change),
              method = "lm",
              span = .01) +
  geom_text(firstnlast, mapping = aes(x = date,
                                      y = percent_change,
                                      label = percent_change),
            vjust = -.5) +
  labs(title = "Working Age Population that are Employed",
       subtitle = "with Best-fit line",
       x = "Date",
       y = "Thousands of People")

grid.arrange(g4, g5, nrow = 2)
```

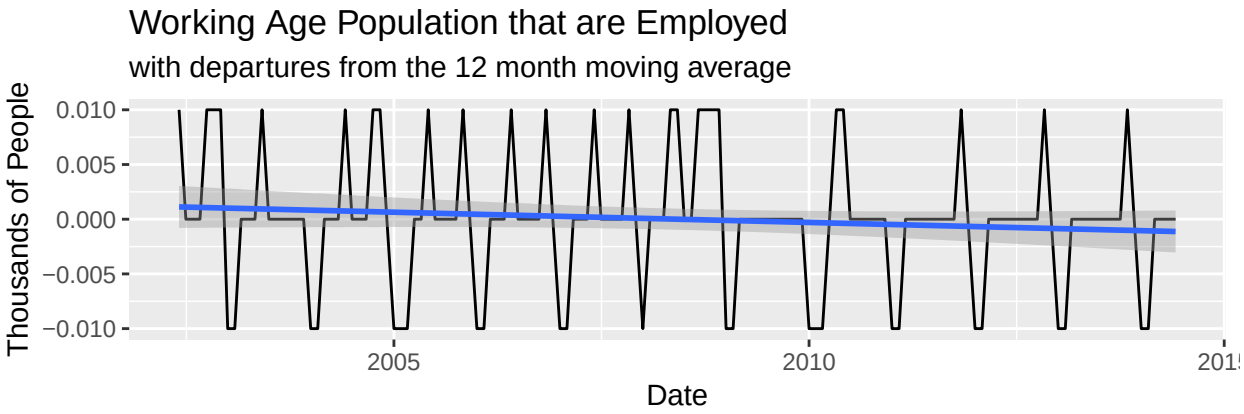


Now we will compute a moving average using `zoo::rollmean()`.

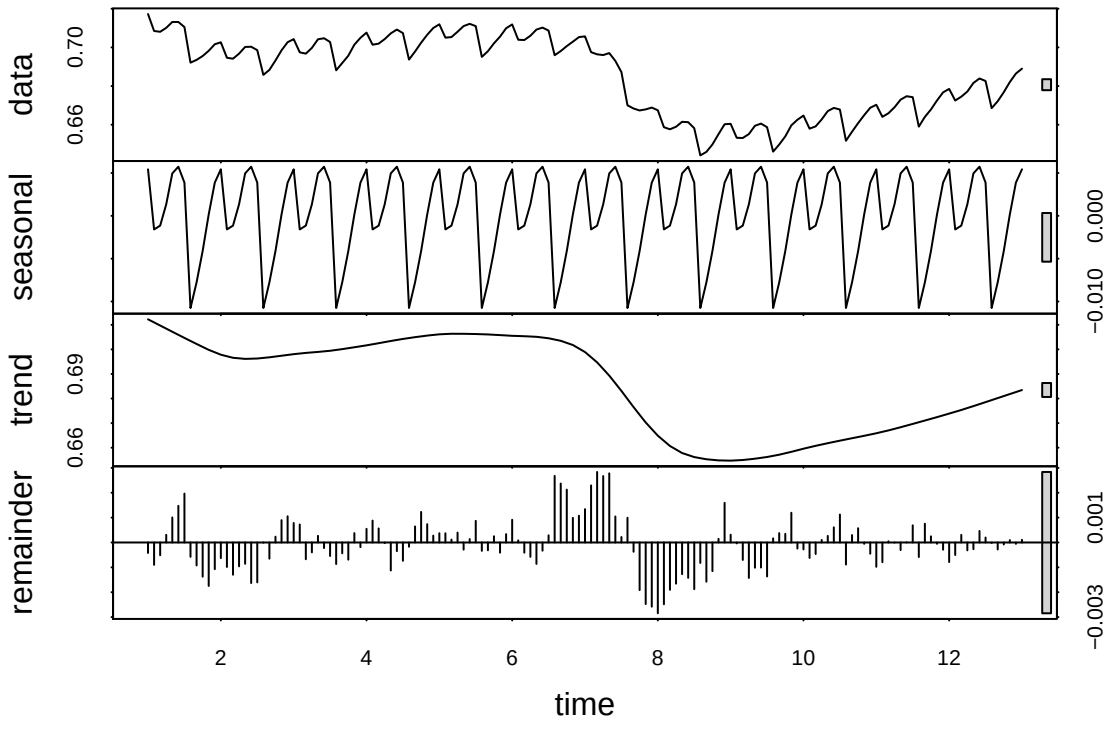
```
pop_emp_recent %>%
  mutate(
    percent_ma = rollmean(percent_emp, 12, fill = NA)
  ) %>%
  filter(!is.na(percent_ma)) -> pop_emp_recent

g6 <- ggplot() +
```

```
  geom_line(pop_emp_recent, mapping = aes(x = date, y = percent_ma)) +
  geom_smooth(pop_emp_recent,
              mapping = aes(x = date, y = percent_ma),
              method = "lm",
              span = .01) +
  labs(title = "Working Age Population that are Employed",
       subtitle = "12 month moving average shown",
       x = "Date",
       y = "Thousands of People")

pop_emp_recent %>%
  filter(!is.na(percent_ma)) %>%
  mutate(rolling_percent_change = round(percent_emp - percent_ma, 2)) -> pop_emp_recent_rolling

g7 <- ggplot() +
  geom_line(pop_emp_recent_rolling, mapping = aes(x = date, y = rolling_percent_change)) +
  geom_smooth(pop_emp_recent_rolling,
              mapping = aes(x = date, y = rolling_percent_change),
              method = "lm",
              span = .01) +
  labs(title = "Working Age Population that are Employed",
       subtitle = "with departures from the 12 month moving average",
       x = "Date",
       y = "Thousands of People")

grid.arrange(g6, g7, nrow = 2)
```



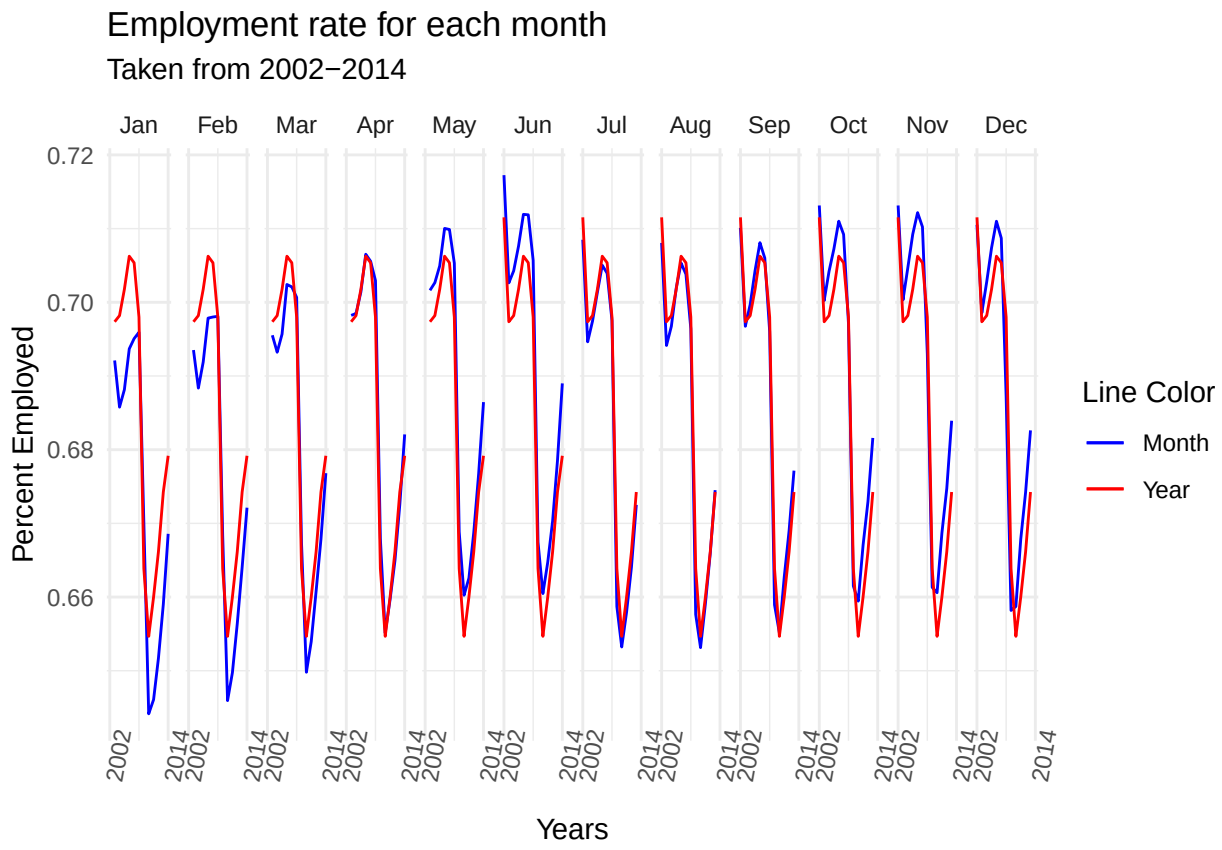



```
plot(stl(x = ts(pop_emp_recent$percent_emp, freq = 12),
     s.window = "periodic", na.action = na.omit))
```




```
pop_emp_recent <- pop_emp_recent %>%
  mutate(month = lubridate::month(date, label = TRUE),
         year = lubridate::year(date))


pop_emp_recent %>%
  group_by(month) %>%
  mutate(
    monthly_avg = mean(percent_emp)
  ) %>%
  select(date, month, year, percent_emp) -> pop_emp_monthly

pop_emp_recent %>%
  group_by(year) %>%
  mutate(
    year_avg = mean(percent_emp)
  ) %>%
  select(date, month, year, year_avg) -> pop_emp_yearly


g8 <- ggplot() +
  geom_line(pop_emp_recent,
            mapping = aes(x = year, y = percent_emp,
            color = "Month")) +
  geom_line(pop_emp_yearly,
            mapping = aes(x = year, y = year_avg,
```

```
            color = "Year")) +
  facet_wrap(~ month, ncol = 12) +
  theme_minimal() +
  scale_x_continuous(breaks = c(2002, 2014)) +
  scale_color_manual(name = "Line Color",
                     values = c(Month= "blue", Year = "red")) +
  labs(title = "Employment rate for each month",
       subtitle = "Taken from 2002-2014",
       y = "Percent Employed",
       x = "Years") +
  theme(axis.text.x = element_text(angle = 80))

g8
```



The final plot asks us to take a look at the difference between each month's score and the corresponding year's average. The first step will be calculating the averages for each year and for each month. Then we will use a `group_by` to calculate the difference between each month over the years.

```
pop_emp_recent %>%
  group_by(year) %>%
  mutate(
    year_avg = mean(percent_emp)
  ) %>%
  ungroup() %>%
  mutate(
    month_to_year_diff = percent_emp - year_avg
```

```
) %>%
ggplot() +
geom_line(mapping = aes(x = year, y = month_to_year_diff)) +
geom_hline(yintercept = 0, linetype = 2) +
facet_wrap(~ month, ncol = 12) +
labs(
  title = "Employment rate for each month compared to average of that year",
  subtitle = "Data from 2002-2014",
  y = "Month-Year difference",
  x = "Years"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 80))
```