BMI 881 Homework 1

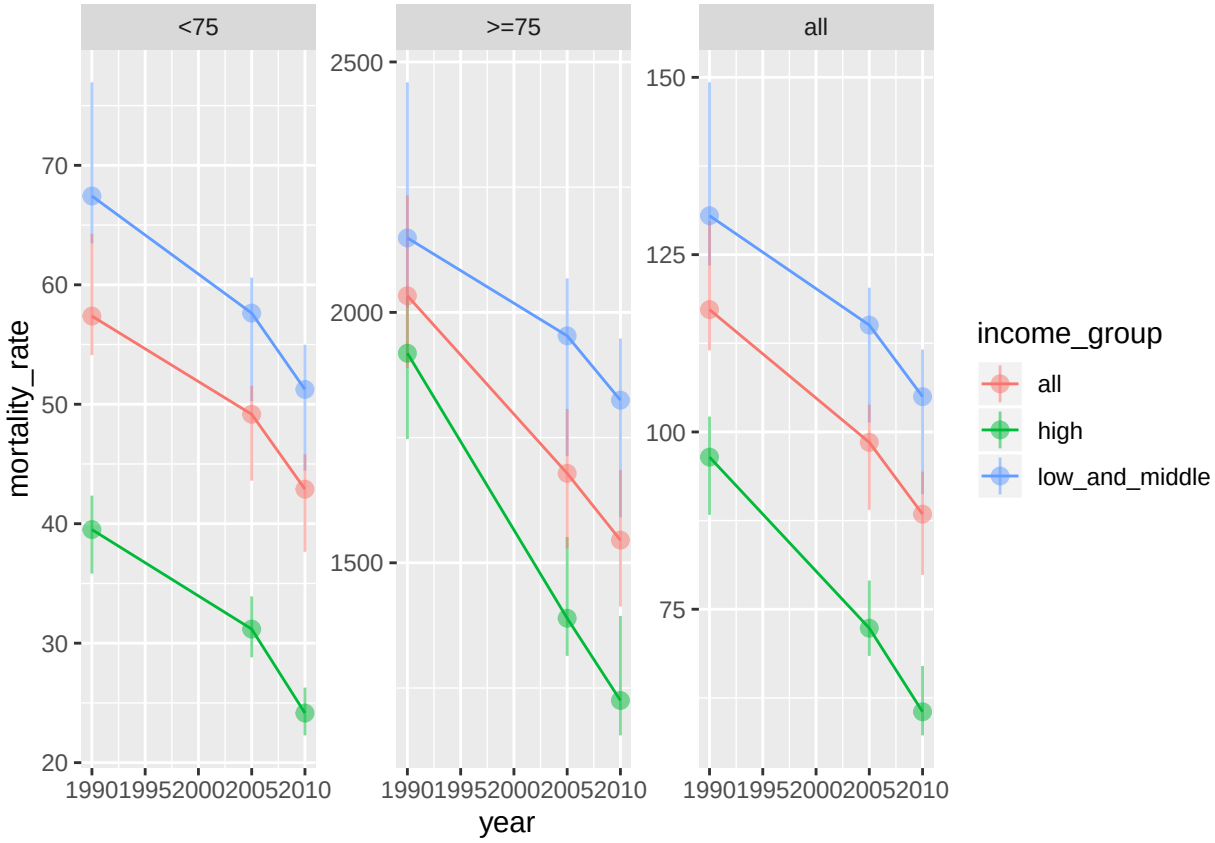
Li Ge

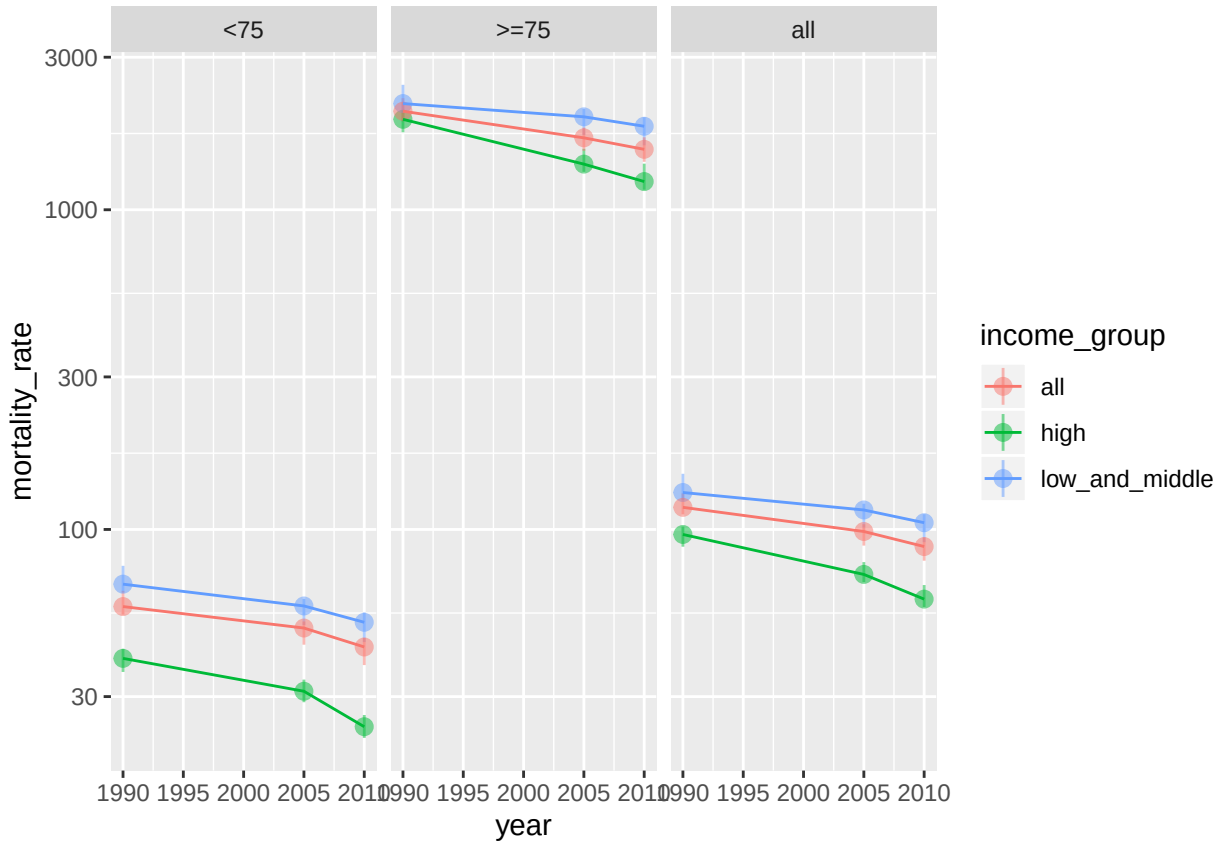
2019-10-09

This is the link to the homework.

Design Choices

- Apparently, the data can be divided into two parts. One part contains the *year, age_group, income_group*, the other part contains the *mortality_rate, interval_low, interval_high*. I should find a way to arrange these different groups. And I should always present the mortality rate together with its interval. ggplot2's facet_grip and geom_pointrange seem to be the perfect fit for it.
- I was originally planning to design a facet_grid with *income_group v. age_group*, each subplot is the *mortality v. year*. But then I realized it would be difficult to compare the mortality rate between different income groups. So I decided to use color to indicate different income groups. And it was so much better.
- Because the mortality rates are drastically different between *>=75 v. rest*. I made 2 figures. For the first one, I used different y-axis for each subplot, which makes the range more obvious. I log-scaled the y-axis for the second graph for easier comparison across all age groups.





Conclusions

- *>=75* age group has the highest mortality rate.
- The higher the income, the lower the mortality rate.
- The mortality rate decreased over the years.

Code

```
knitr::opts_chunk$set(
  echo = FALSE,
  fig.align = "center"
)

library(ggplot2)
# read the data
stroke <- read.csv("https://kbroman.org/BMI881/assets/feigin2014_table1_mortality.csv")

# plot
ggplot(data = stroke, mapping = aes(x = year, y = mortality_rate, color = income_group)) +
  facet_wrap(. ~ age_group, scale = "free_y") +
  geom_pointrange(aes(ymin = interval_low, ymax = interval_high), alpha = 0.5) +
  geom_line()
```

```r
ggplot(data = stroke, mapping = aes(x = year, y = mortality_rate, color = income_group)) +
  facet_wrap(. ~ age_group) +
  scale_y_log10() +
  geom_pointrange(aes(ymin = interval_low, ymax = interval_high), alpha = 0.5) +
  geom_line()

# this R markdown chunk generates a code appendix
```