Problem Set 1

Bern DySart

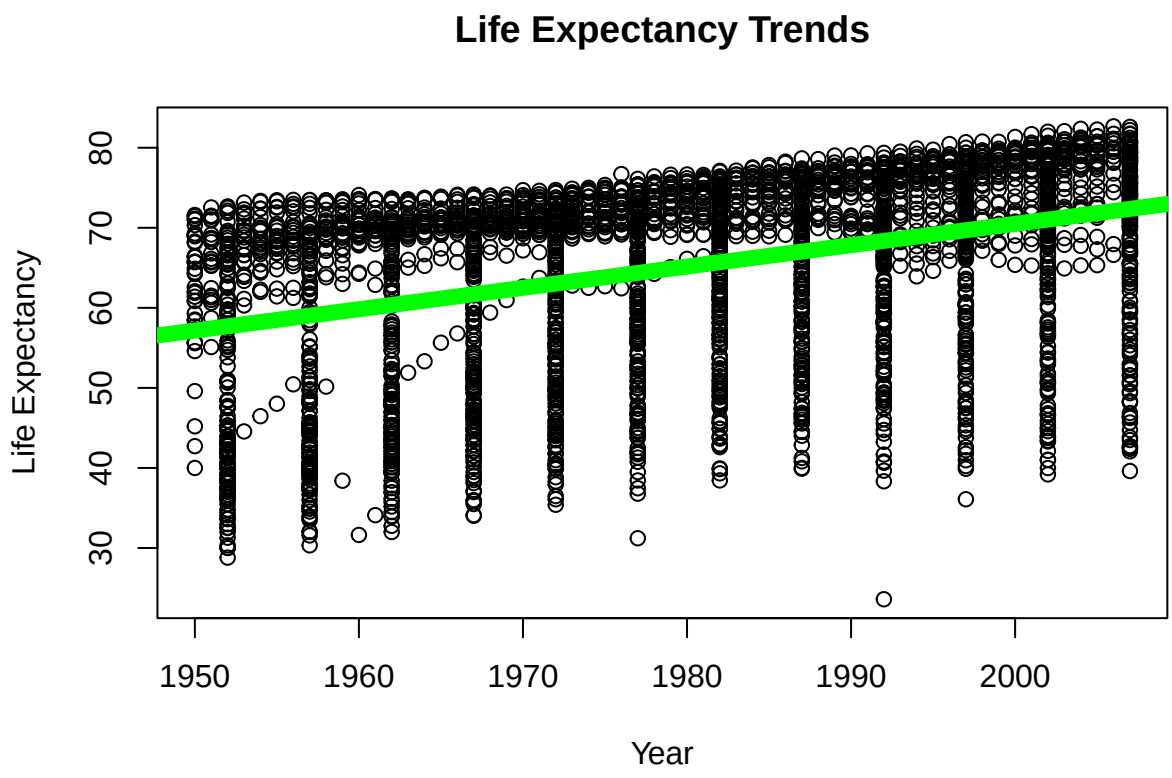
2024-01-24

1.There are 3,313 observations in the dataset. There are 6 variables in the dataset.gapminder_unfilterd contains data from years 1950 - 2007. Every country is covered in the data set for each year.

```
gapminder = gapminder_unfiltered
```

2. See plot below

```
plot(lifeExp~year, data = gapminder,
     xlab = 'Year',
     ylab = 'Life Expectancy',
     main = 'Life Expectancy Trends')
model1 =
  lm(lifeExp ~ year, data = gapminder)
abline(model1,col = 'green', lw = 8)
```



3. See table below

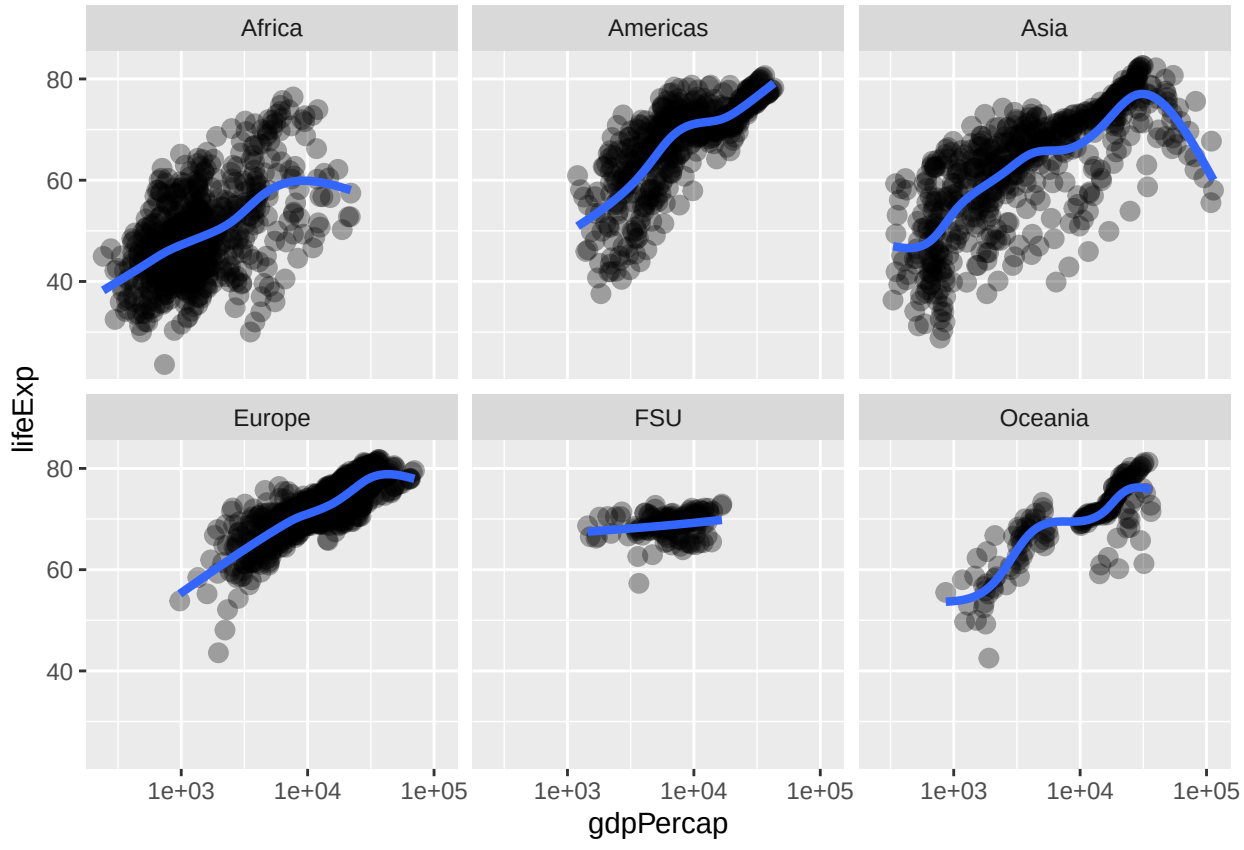
```
outcome =
  count(gapminder, continent)
kable(outcome)
```

continent	n
Africa	637
Americas	470
Asia	578
Europe	1302
FSU	139
Oceania	187

4. See plot below

```
p <-ggplot(gapminder,
           aes(x = gdpPercap, y = lifeExp))
p <- p + scale_x_log10()
p +
  geom_point(alpha = (1/3), size = 3) +
  facet_wrap(~ continent) +
  geom_smooth(lwd = 1.5, se = FALSE)
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



5. See fixed code

```
flights |>
  filter(dest=="IAH") |>
  group_by(year,month,day) |>
  summarize(
    n=n(),
    delay=mean(arr_delay, na.rm=TRUE)
  ) |>
  filter(n>10)

flights |>
  filter(carrier=="UA",
         dest%in%c("IAH", "HOU"),
         sched_dep_time > 0900,sched_arr_time<2000) |>
  group_by(flight) |>
  summarize(
    delay=mean(arr_delay,na.rm=TRUE),
    cancelled=sum(is.na(arr_delay)),n=n()
  ) |>
  filter(n>10)
```