

STA 326 2.0 Programming and Data Analysis with R

Functionals

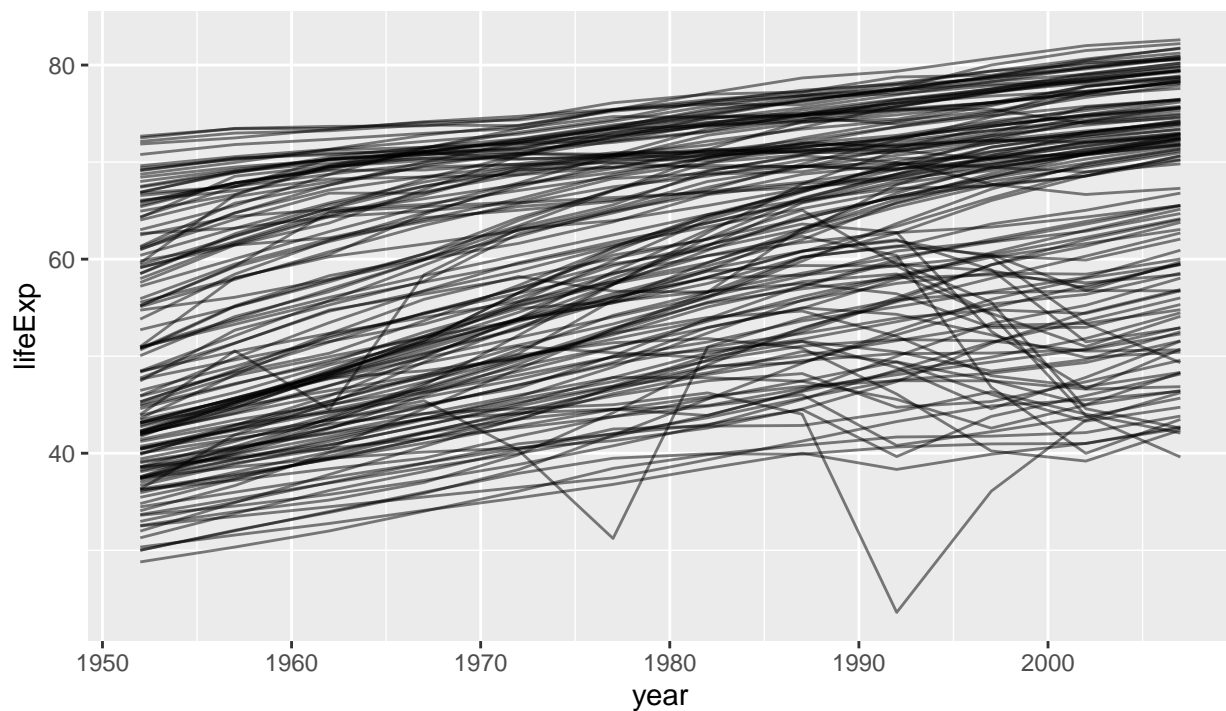
Question:

Use the gapminder dataset.

```
library(gapminder)
data(gapminder)
head(gapminder)
```

```
# A tibble: 6 x 6
  country    continent  year  lifeExp    pop  gdpPercap
  <fct>      <fct>    <int>  <dbl>    <int>  <dbl>
1 Afghanistan Asia      1952   28.8  8425333   779.
2 Afghanistan Asia      1957   30.3  9240934   821.
3 Afghanistan Asia      1962   32.0 10267083   853.
4 Afghanistan Asia      1967   34.0 11537966   836.
5 Afghanistan Asia      1972   36.1 13079460   740.
6 Afghanistan Asia      1977   38.4 14880372   786.
```

```
library(ggplot2)
ggplot(gapminder, aes(x=year, y=lifeExp, group=country)) +
  geom_line(alpha=0.5)
```



Write an R code to estimate β_0 and β_1 of $lifeExp = \beta_0 + \beta_1 year$ using least squares approach for each country.

Plot the distribution of $\hat{\beta}_1$ values. Interpret the results.

```
library(purrr)
gapminder_countries <- split(gapminder, gapminder$country)
estimated.coef <- gapminder_countries %>%
  map(~lm(lifeExp ~ year,
         data = .x)) %>%
  map_df(coef)
head(estimated.coef)
```

```
# A tibble: 6 x 2
  `(Intercept)` year
  <dbl> <dbl>
1    -508. 0.275
2    -594. 0.335
3   -1068. 0.569
4    -377. 0.209
5    -390. 0.232
6    -376. 0.228
```

```
ggplot(estimated.coef, aes(x=year)) + geom_histogram(col="white")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

