R Notebook – gapminder data exploration

Neil Yetz

Contents

The 3 dashes above (In the code) indicate that we are woirking within the Header information of out R notebok

Load Packages for this session

```r
library(tidyverse)
```

Import data

The gapminder dataset is provided by an organization called Gapminder.org. It contains data on 142 countries. For each country the dataset provides values for life expectancy, GDP per capita, and population - with these data available every five years from 1952 to 2007.

```r
#notes work in code chunks like normal comments

gm <- read_csv(file="gapminder.csv")
```

Summarize the variables

```r
summary(gm)
```
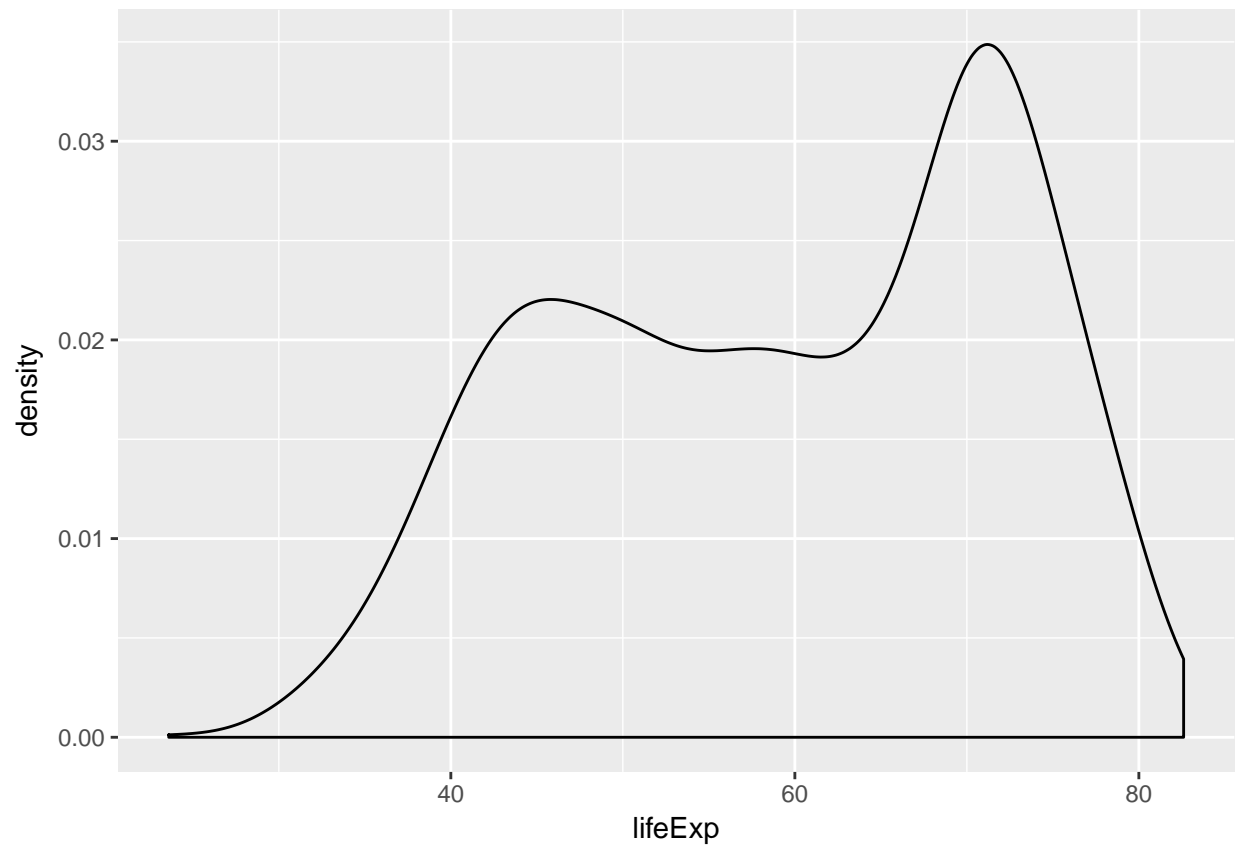
```
##    country           continent              year          lifeExp
##  Length:1704        Length:1704        Min.   :1952    Min.   :23.60
##  Class :character   Class :character   1st Qu.:1966    1st Qu.:48.20
##  Mode  :character   Mode  :character   Median :1980    Median :60.71
##                                        Mean   :1980    Mean   :59.47
##                                        3rd Qu.:1993    3rd Qu.:70.85
##                                        Max.   :2007    Max.   :82.60
##       pop              gdpPercap
##  Min.   :6.001e+04   Min.   :   241.2
##  1st Qu.:2.794e+06   1st Qu.:  1202.1
##  Median :7.024e+06   Median :  3531.8
```

```
## Mean    :2.960e+07   Mean    : 7215.3
## 3rd Qu.:1.959e+07   3rd Qu.: 9325.5
## Max.    :1.319e+09   Max.    :113523.1
```

Create a plot

```
ggplot(gm, aes(x=lifeExp)) + geom_density()
```



```
ggplot(gm, aes(x = lifeExp)) +
geom_density(fill = "yellow") +
  labs(title = "What is the density of life expectancy in the Gapminder dataset?",
x = "Life expectancy in years", y = "Density")
```

What is the density of life expectancy in the Gapminder dataset?

Enhancement to the plot

```
ggplot(gm, aes(x = lifeExp, group = continent, fill=continent)) +
geom_density(alpha=.2) +
labs(title = "Does the distribution of life expectancy differ by continent?",
x = "Life expectancy in years", y = "Density")
```

Does the distribution of life expectancy differ by continent?