Microbiol612

Corine Jackman

March 18, 2016

Introduction:

In order to become familiar with R, we analyzed data from the Gapminder file which contained information for different countries regarding the year, population, continent, life expecancy, and GDP per capita. The data was saved to myprojects and was uploaded during class using the following 2 lines of code shown below.

```
getwd()
```

```
## [1] "/Users/corinemj/myproject"
```

```
gapminder<-read.table("./r-novice-gapminder-files/data/gapminder-FiveYearData.csv", s
ep=',', header=T)
head(gapminder)
```

```
##       country year      pop continent lifeExp gdpPercap
## 1 Afghanistan 1952  8425333      Asia  28.801  779.4453
## 2 Afghanistan 1957  9240934      Asia  30.332  820.8530
## 3 Afghanistan 1962 10267083      Asia  31.997  853.1007
## 4 Afghanistan 1967 11537966      Asia  34.020  836.1971
## 5 Afghanistan 1972 13079460      Asia  36.088  739.9811
## 6 Afghanistan 1977 14880372      Asia  38.438  786.1134
```

Assignment:

- **Part 1:** What is the mean and standard deviation of the life expectancies for each of the continents?
- **Part 2:** Plot the mean and standard deviation for the continents.

Part 1:

In order to do this, the dplyr package was installed, loaded and the group_by() and summarize() functions were used.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

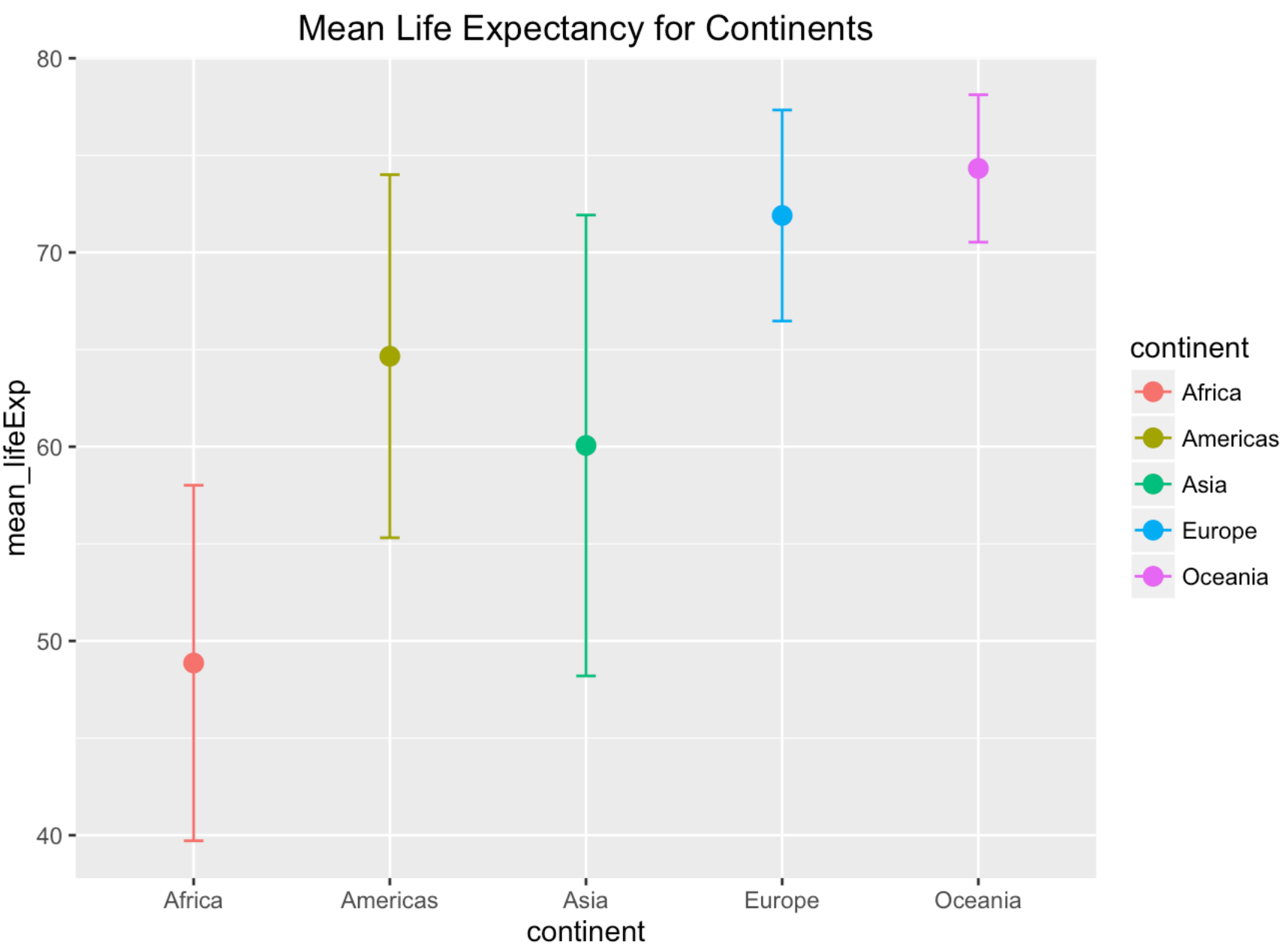
```
lifeExp_bycontinents <- gapminder %>%
    group_by(continent) %>%
    summarize(mean_lifeExp=mean(lifeExp),
    sd_lifeExp=sd(lifeExp))

lifeExp_bycontinents
```

```
## Source: local data frame [5 x 3]
##
##   continent mean_lifeExp sd_lifeExp
##      (fctr)        (dbl)      (dbl)
## 1    Africa     48.86533   9.150210
## 2  Americas     64.65874   9.345088
## 3      Asia     60.06490  11.864532
## 4    Europe     71.90369   5.433178
## 5   Oceania     74.32621   3.795611
```

Part 2:

In order to plot this data, ggplot was installed by selecting “Packages” in Rstudio and clicking “ggplot2”. Then, the following code shows the x axis, y axis and title as “continent”, “mean life expectancy” and “Mean Life Expectancy for Continents”, respectively. The “color=continent” designates a color for each continent, the geom_errorbar() designates error bars and geom_point() shows the mean life expectancy as an individual point on the plot. Africa has the lowest mean life expectancy. Oceania has the highest mean life expectancy.

```
library(ggplot2)
ggplot(lifeExp_bycontinents, aes(x = continent, y = mean_lifeExp, color=continent)) +
ggtitle("Mean Life Expectancy for Continents") +
  geom_errorbar(aes(ymin=mean_lifeExp-sd_lifeExp, ymax=mean_lifeExp+sd_lifeExp), widt
h=0.1) +
  geom_point(size=3)
```


Mean Life Expectancy for Continents
mean_lifeExp
80
70
60
50
40
Africa
Americas
Asia
Europe
Oceania
continent
Africa
Americas
Asia
Europe
Oceania