Homework 1: a Gapminder Exercise

Alexander Chang

February 7, 2018

Homework 01 Exercise - Task 1

Modify the sentence above to also provide the standard deviation, median and sample size for life expectancy, set `digits=2`.

Hint: Read help pages for the functions `sd()`, `median()`, and `length()`.

Answer Task 1

The mean life expectancy is 59.47 years with a standard deviation of 12.92. The median life expectancy is 60.71 and the sample size is 1704,

Look at a statistic by continent

Using the 1st code example at the `gapminder` Github repository at https://github.com/jennybc/gapminder, use the `aggregate()` command to see the median life expectancy by continent. You'll notice that the 1st variable listed is the "lifeExp" variable we want run "by" "continent". The "by" is indicated using the *tilde* symbol ~. The 2nd variable listed is "continent" - this 2nd variable is usually a "factor-type" variable or group variable. *Hint: Try running lifeExp by year to get median lifeExp for each year.*

```
aggregate(lifeExp ~ continent, gapminder, median)

##   continent lifeExp
## 1    Africa 47.7920
## 2  Americas 67.0480
## 3      Asia 61.7915
## 4    Europe 72.2410
## 5   Oceania 73.6650
```

Homework 01 Exercise - Task 2

Modify the r code chunk above to also provide the mean and standard deviation for life expectancy by continent.

Task 2 Answer

```
#Mean
aggregate(lifeExp ~ continent,gapminder,mean)

##   continent  lifeExp
## 1    Africa 48.86533
## 2  Americas 64.65874
## 3      Asia 60.06490
## 4    Europe 71.90369
## 5   Oceania 74.32621
```

```
#standard Deviation
aggregate(lifeExp ~ continent,gapminder,sd)
```

```
##   continent   lifeExp
## 1    Africa  9.150210
## 2  Americas  9.345088
## 3      Asia 11.864532
## 4    Europe  5.433178
## 5   Oceania  3.795611
```

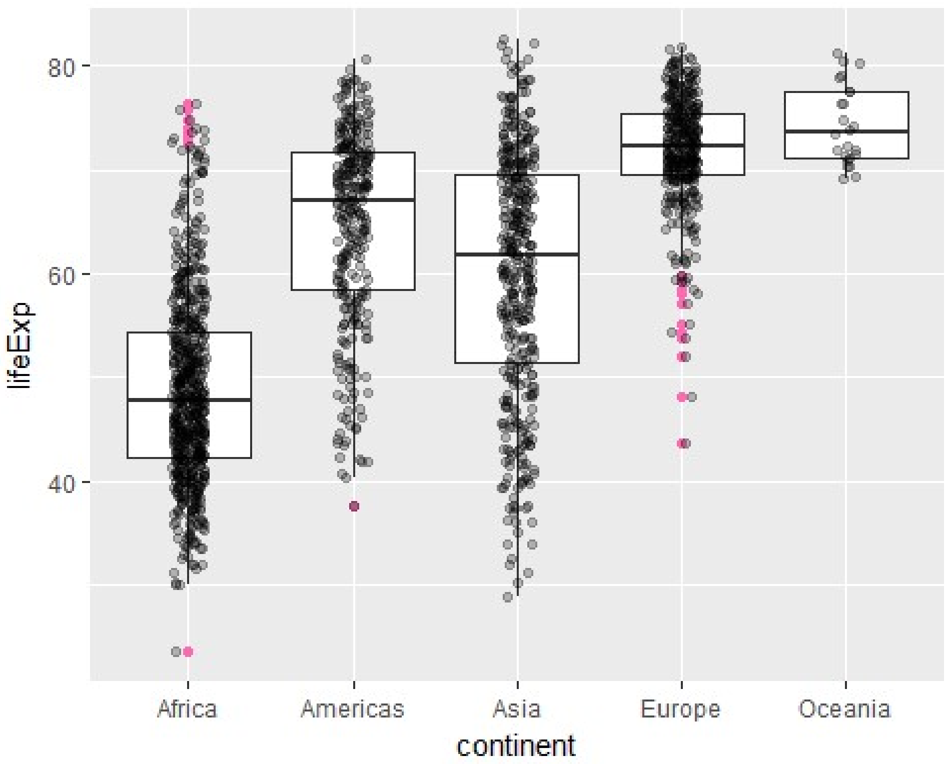
Plot Life Expectancy by Continent

For the following plot, we will use the `ggplot2` package. Please be sure to install the package if you haven't already. Go to RStudio "Tools" menu and select "Install Packages" and type in "ggplot2". The R code chunk below will load the `ggplot2` package using the `library(ggplot2)` command. For now, we are using the code provided at the **Gapminder** Github repository located at https://github.com/jennybc/gapminder. We will explore and explain the ggplot syntax in further detail later this semester. For now, notice that in the `aes()` section of code below, there are 2 "aesthetics" provided to `ggplot()` - namely that `x = continent` and `y = lifeExp`. The code below generates 2 plots overlaid on top of each other. The 1st is a boxplot generated by `geom_boxplot` and then a 2nd plot of "jittered" points is overlaid using the `geom_jitter()` command.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```
ggplot(gapminder, aes(x = continent, y = lifeExp)) +
  geom_boxplot(outlier.colour = "hotpink") +
  geom_jitter(position = position_jitter(width = 0.1, height = 0), alpha =
1/4)
```

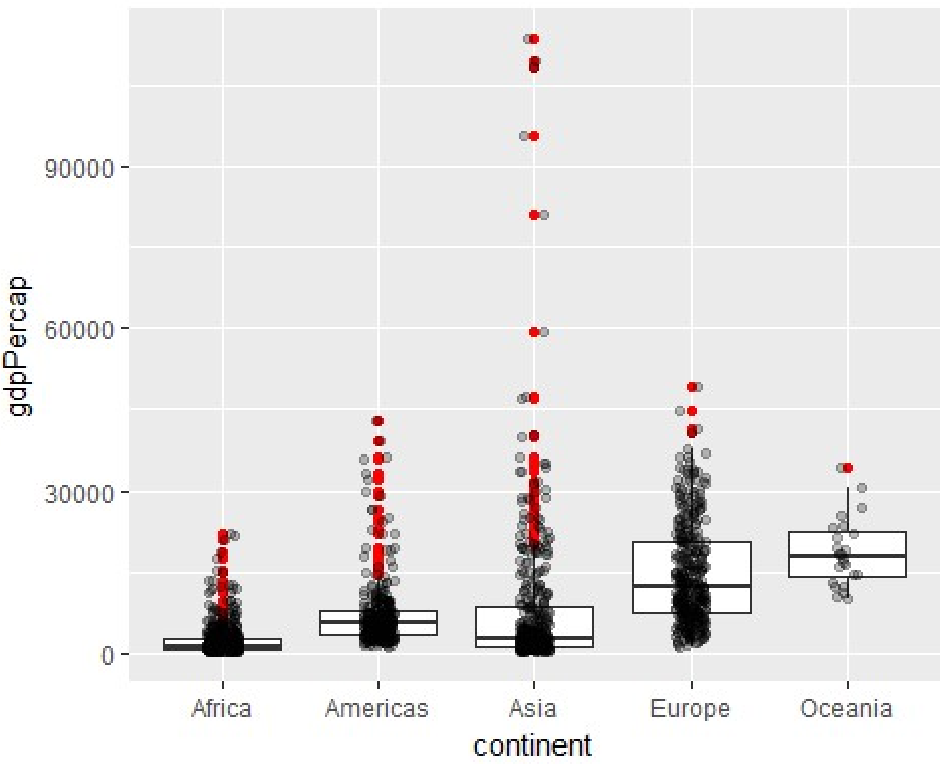


Homework 01 Exercise - Task 3

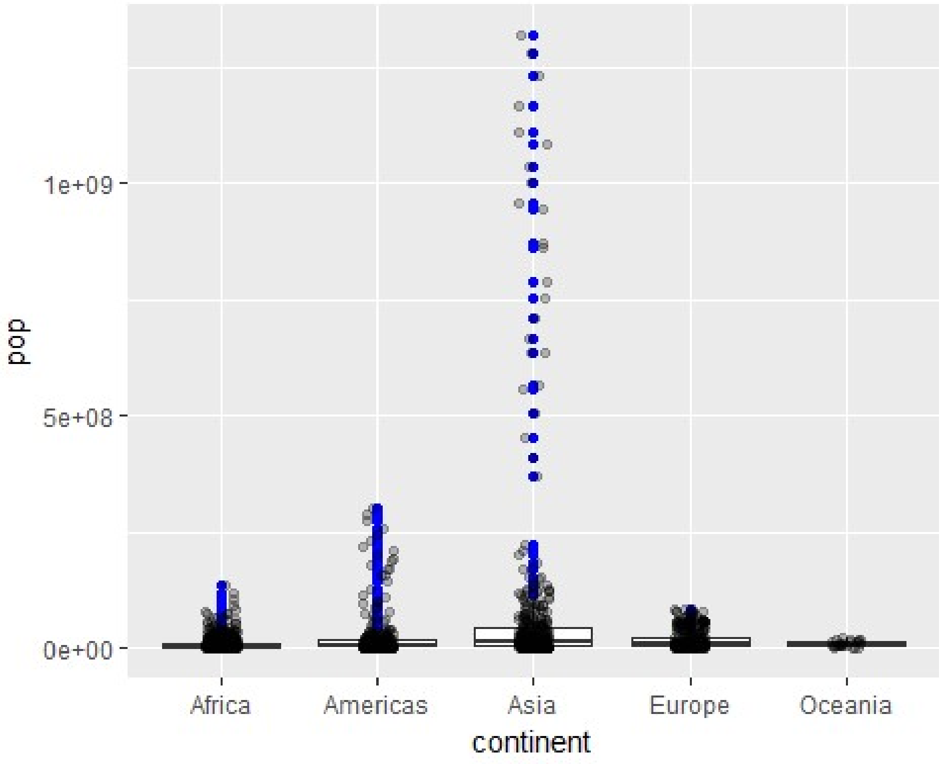
Modify the R code chunk above to now produce two more plots: one for `gdpPercap` instead of `lifeExp` and one for pop instead of `lifeExp`. *Hint: change out the variable for y above. You will need the entire code chunk run twice - once for each different plot.*

Task 3 Answer

```
library(ggplot2)
ggplot(gapminder, aes(x = continent, y=gdpPercap)) +
  geom_boxplot(outlier.color = 'red') +
  geom_jitter(position = position_jitter(width = 0.1,height = 0),alpha=1/4)
```



```r
ggplot(gapminder, aes(x = continent, y=pop)) +
  geom_boxplot(outlier.color = 'blue') +
  geom_jitter(position = position_jitter(width = 0.1,height = 0),alpha=1/4)
```

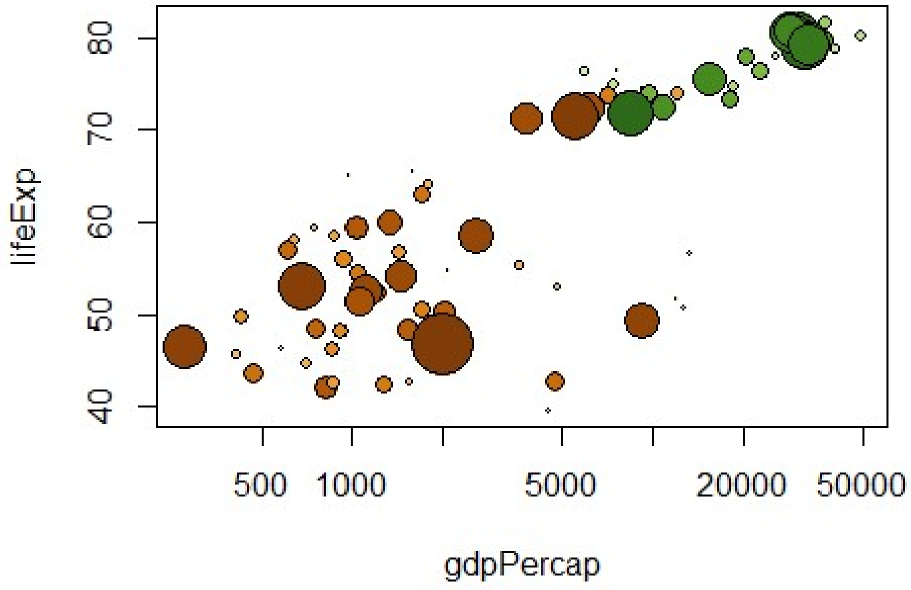


Selecting a Data subset and Customizing Plot Colors

For this next example, we will use the built-in "country_colors" dataset defined in the gapminder package to create an updated data frame (cope of the gapminder dataset) called gap_with_colors. *NOTE: To see the built-in datasets in the gapminder package, type data(package = "gapminder") in the console.* After setting up the custom colors, the next line of R code defines which continents we plan to "keep" in our plot. This defines and object called keepers below. In this case, we are "keeping" Africa and Europe and the year 2007. So, we are subsetting the dataset for 2 countries and 1 year. Then the plot() command creates the scatterplot of life expectancy by GDP per capita with a dot representing each county within 2 continents (Africa and Europe) for only 2007. The size of each dot is represents a 3rd variable "pop" for population size.

```r
# for convenience, integrate the country colors into the data.frame
gap_with_colors <-
  data.frame(gapminder,
             cc = I(country_colors[match(gapminder$country,
                                         names(country_colors))]))

# bubble plot, focus just on Africa and Europe in 2007
keepers <- with(gap_with_colors,
                continent %in% c("Africa", "Europe") & year == 2007)
plot(lifeExp ~ gdpPercap, gap_with_colors,
     subset = keepers, log = "x", pch = 21,
     cex = sqrt(gap_with_colors$pop[keepers]/pi)/1500,
     bg = gap_with_colors$cc[keepers])
```

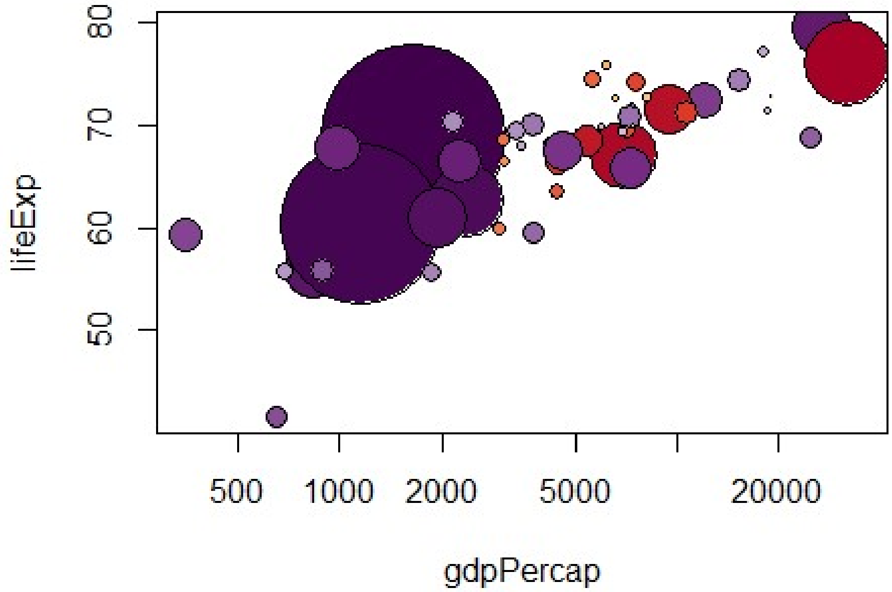


Homework 01 Exercise - Task 4

Modify the code chunk above to make a bubble plot for 2 other continents: Americas and Asia for the year 1992. *Hint: modify the variables listed in the "keepers" line of code for the 2 countries and the year.* ### **Task 4 Answers**

```
gap_with_colors <-
  data.frame(gapminder,
             cc = I(country_colors[match(gapminder$country,
                                         names(country_colors))]))

# bubble plot, with focus on Americas and Asia for the year 1992
keepers <- with(gap_with_colors,
                continent %in% c("Americas", "Asia") & year == 1992)
plot(lifeExp ~ gdpPercap, gap_with_colors,
     subset = keepers, log = "x", pch = 21,
     cex = sqrt(gap_with_colors$pop[keepers]/pi)/1500,
     bg = gap_with_colors$cc[keepers])
```



Homework 01 Exercise - Task 5

Task 5 link

[https://github.com/theamazingchang/N741/blob/master/Homework1_Chang.Rmd][https://github.com/theamazingchang/N741/blob/master/Homework1_Chang.html]

- For your report, modify the YAML to put a good TITLE for your report and put in YOUR NAME and change the DATE.
- If you have MiKTeX or MacTex installed and running on your computer, you can compile the report directly to PDF (“Knit to PDF”). However, if you didn’t install these, you should still be able to “Knit to HTML” or “Knit to Word”. Once you’ve got the HTML or Microsoft Word DOCX file created, you can print these to PDF.
- Make sure your report completes the 4 tasks above with your associated R chunks along with a description or commentary on what you did and the result.
- Submit the PDF homework using Canvas for N741.
- Include in your report a link to your Github repository with your RMD file and any other associated output you created. In the body of your Rmarkdown (RMD) report use the `[]()` syntax to embed the URL “web link”. For example, the link for this exercise is located at https://github.com/melindahiggins2000/N741gapminder1 which was typed in to the Rmarkdown file as `[https://github.com/melindahiggins2000/N741gapminder1](https://github.com/melindahiggins2000/N741gapminder1).`