# ECON 329 R Bootcamp Solutions

*Matt Kaye*

*4/3/2019*

```
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE)

#put message = F and warning = F in this chunk
#so that you don't print warning or package startup messages
```

```
check.packages <- function(pkg){
    new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
    if (length(new.pkg))
        install.packages(new.pkg, dependencies = TRUE)
    sapply(pkg, require, character.only = TRUE)
}
#install if you don't have these
#this chunk is only here for install help -- don't put install chunks in your RMD
#use the console to install the packages instead
check.packages("tidyverse")
check.packages("gapminder")
check.packages("stargazer")
check.packages("kableExtra")
```

## This is how you make a header – more # signs will make the header smaller

```
#empty working environment to keep R workspace clean
rm(list = ls())

#load packages
library(gapminder)
library(tidyverse)
library(stargazer)
library(knitr)

#sets seed for random number generator so your bootstraps, random samples, etc
#will always knit the same
set.seed(10065)
```

## Question 2 and Inline R!

```
#need results = 'asis' for stargazer(type = "latex") to work

df <- gapminder

df %>% #pipe operator -- makes df (left side) the first parameter in data.frame()
```

```r
  data.frame() %>%
  select(4:6) %>% #use select() from dplyr to pick columns 4 through 6
  stargazer(header = FALSE, font.size = 'footnotesize') #stargazer summary stats table
```

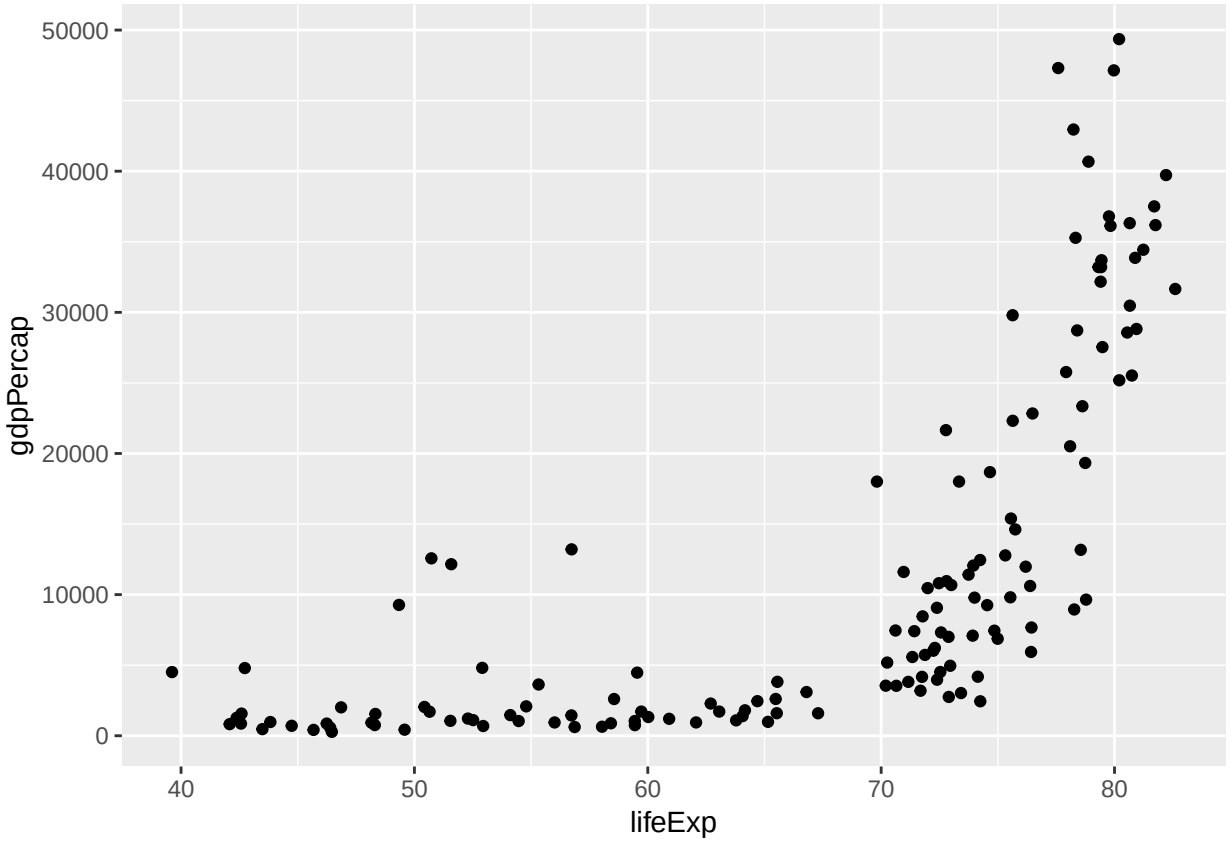
Table 1:

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| lifeExp | 1,704 | 59.474 | 12.917 | 23.599 | 48.198 | 70.846 | 82.603 |
| pop | 1,704 | 29,601,212.000 | 106,157,897.000 | 60,011 | 2,793,664 | 19,585,222.0 | 1,318,683,096 |
| gdpPercap | 1,704 | 7,215.327 | 9,857.455 | 241.166 | 1,202.060 | 9,325.462 | 113,523.100 |

You should type your answers to questions here, like this. If you have something in an R chunk that you want to print out as part of an answer (like the mean of gdpPercap, for example), you can print it inline like this: 7215.327. This will come in especially handy if you mess something up in question 1 that causes all of your answers to later questions to be off... Instead of needing to change everything, inline R will fix update everything for you.

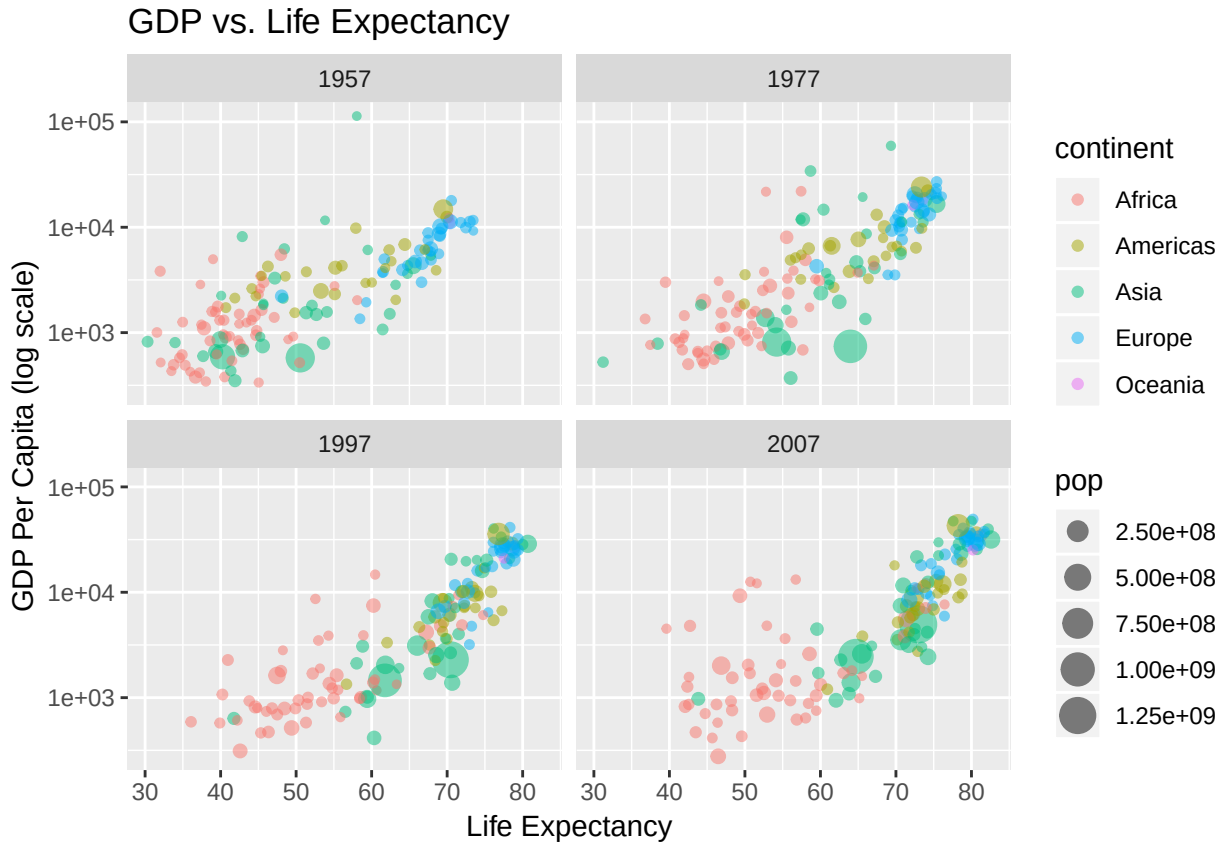
## Question 3

```r
df %>% #use the pipe operator -- passes df into filter as first parameter
  filter(year == 2007) %>% #subset data to only 2007
  ggplot(aes(x = lifeExp, y = gdpPercap))+ #create ggplot
  geom_point() #add layer for point (i.e. scatter plot)
```



## Question 4

```
df %>%
  #use %in% with filter to keep multiple years
  filter(year %in% c(1957, 1977, 1997, 2007)) %>%
  #add colors and sizes
  ggplot(aes(x = lifeExp, y = gdpPercap, color = continent, size = pop))+
   #alpha blend the points -- make them translucent to make the plot clearer
  geom_point(alpha = .5)+
  scale_y_log10()+ #put the y on log scale
  #facet wrapper -- makes one plot for each category of the variable after "~"
  facet_wrap(~year, ncol = 2, nrow = 2)+
  #add plot labels
  labs(x = "Life Expectancy",
       y = "GDP Per Capita (log scale)",
       title = "GDP vs. Life Expectancy")
```



## Question 5

```
df %>%
  filter(year %in% c(1957, 2007)) %>%
  group_by(continent, year) %>% #sets continent and year as "grouping variables"
                                #lets us perform operations on every row within each group
  summarise(gdp = mean(gdpPercap)) %>% #calc mean gdpPercap for each continent-year pair
  spread(year, gdp) %>%  #create two new columns from the data frame we created above
  kable(format = 'latex', digits = 0) %>% #use kable to create nicely formatted tables
  kableExtra::row_spec(0, bold = TRUE) %>% #made header bold
  kableExtra::kable_styling(latex_options = "striped") %>% #made table striped
  kableExtra::add_header_above("GDP Per Capita")
```

| GDP Per Capita | | |
|---|---|---|
| **continent** | **1957** | **2007** |
| Africa | 1385 | 3089 |
| Americas | 4616 | 11003 |
| Asia | 5788 | 12473 |
| Europe | 6963 | 25054 |
| Oceania | 11599 | 29810 |

## Question 6:

```
codes <- country_codes
df <- df %>%
  #left join attaches all rows matching the "by" in table 2 to table 1
  left_join(codes, by = "country") %>%
  select(2, 1, 7, 3:6) %>% #selects columns by index
  mutate(pop = pop / 1000) #adds column

df %>%
  filter(year == 2007) %>%
  sample_n(10, FALSE) %>% #random sample of 10 rows without replacement
  arrange(continent, country) %>%
  kable(format = 'latex', digits = 0) %>% #use kable to create nicely formatted tables
  kableExtra::row_spec(0, bold = TRUE) %>% #made header bold
  kableExtra::kable_styling(latex_options = "striped") %>% #made table striped
  kableExtra::add_header_above("Random Country Sample") %>% #add title
  kableExtra::footnote(symbol = "population in thousands") #add a note
```

| Random Country Sample | | | | | | |
|---|---|---|---|---|---|---|
| **continent** | **country** | **iso_alpha** | **year** | **lifeExp** | **pop** | **gdpPercap** |
| Africa | Malawi | MWI | 2007 | 48 | 13327 | 759 |
| Americas | Costa Rica | CRI | 2007 | 79 | 4134 | 9645 |
| Americas | Ecuador | ECU | 2007 | 75 | 13756 | 6873 |
| Americas | Trinidad and Tobago | TTO | 2007 | 70 | 1057 | 18009 |
| Americas | Venezuela | VEN | 2007 | 74 | 26085 | 11416 |
| Asia | Iran | IRN | 2007 | 71 | 69454 | 11606 |
| Asia | Malaysia | MYS | 2007 | 74 | 24821 | 12452 |
| Asia | West Bank and Gaza | PSE | 2007 | 73 | 4018 | 3025 |
| Europe | Croatia | HRV | 2007 | 76 | 4493 | 14619 |
| Europe | Germany | DEU | 2007 | 79 | 82401 | 32170 |

* population in thousands

## Appendix Of Useful Latex Stuff

If you want to make some text **bold** or *italic*, do this

If you want to make your regression equations look nice, do something like this:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 \cdot X + \epsilon$$

The dollar signs tell Rmarkdown to use LaTeX's math mode, which lets you use Greek letters, subscripts, etc.

lets you indent a line stops a line from being indented.

- lets you add bullet points

1. same idea for adding numbered bullets

if you want to make equations really fancy (to do derivations of properties of summations, for example), you can do something like this:

$$
\begin{aligned}
X &= Y \\
Y &= Z \\
\implies X &= Z
\end{aligned}
$$

this is how you center text

and this is how you skip some space

Now, some potentially helpful side notes: First, Detexify (for LaTeX) and StackOverflow are great resources for help on LaTeX and R questions. Also, the R documentation on all of these functions is really good. Doing something line ?filter or ?ggplot in the R console will open the documentation (or you can just Google it).