

---

title: 'Lab #1: Introduction to R'  
date: "STAT 250 - Spring 2020"  
output:  
html\_document: default

---

## Name: Parth Iyer

### Important package imports

Importing pandas for handling csv's & math for doing Math!

```
import pandas as pd
import math
```

## Skills

---

- Demonstrate basic skills in R coding
- Recognize R syntax to decode new R commands
- Be able to identify basic data types in R and how to encode them (see Primer)
- Be able to create and manipulate vectors in R
- Be able to create and manipulate dataframes in R
- Be able to use variables in arithmetic functions in R

**Please make sure to show all R code and output after each question so that we can see your work.** Write a sentence for each numerical value produced describing its meaning **in context with the proper units**. Be sure to submit your full project as a .zip file that includes your edited, renamed .Rmd file and the .Rproj file to receive credit. Detailed instructions for procedural parts of labs are found in the "General Lab Directions" on your iLearn page; refer to it for how to download/upload/etc. If you don't recall how to do something, you should first refer to your lecture notes, in-class examples, and the primers from Lab #1 and #2. Your lecture notes and readings on these topics are also good resources. This lab should be completed **individually**.

---

For this lab, you should begin by reading the accompanying primer and referring to your notes from in-class tutorials and in-class activities.

## Part 1: Data Organization

---

1.1. Is this block of raw data is organized correctly? Why?

Length (mm)	Sex
145 mm	F
163 mm	F
132 mm	M

Length (mm)	Sex
143 mm	M

Yes, each individual is a row.

1.2. Is this block of raw data organized correctly? Why?

Info	A	B	C	D
Length	145	163	132	143
Sex	F	F	M	M

No, Each individual is not a row.

## Part 2: R Coding

Stephen Curry is one of the most prolific scorers currently in the NBA. We can look at the number of points he scored during games in 2015.

2.1. Read in the csv file `curry2015.csv` and store it as the object `curry` by modifying the code below to fill in each blank. Some blanks in this lab will have hints to the code you need written in the blanks like this: *hint*.

```
curry = pd.read_csv("curry2015.csv")
```

2.2. Find the mean number of points Curry scored in 2015 from the selected games by modifying the code below to replace the blanks with the correct code. (You can use the function `names()` and apply it to your dataframe object to see the names of the variables, if needed.)

```
avg = sum(curry.Point)/len(curry.Point)
```

2.3. Insert a code chunk and write the code to print your raw data as the output. Why do you think you got the answer you did in 2.2?

```
print(avg)
```

The answer in 2.2 is Nan or Not A Number. This is because of blanks in the record.

2.4. We need to add an argument to `mean()` to calculate the mean number of points in games Curry played. The games he didn't play are coded with an `NA`. We need to tell R to ignore them for the sake of the `mean()` function. The argument we need to use to do that is called `na.rm`. Modify the code below to name the second argument. What is the mean number of points Curry scores per game in which he plays?

```
cleanedset = curry.values.tolist()
for i in range(len(cleanedset)):
    if math.isnan(cleanedset[i][1]):
        cleanedset[i][1] = 0
curry = pd.DataFrame(cleanedset, columns=["Game", "Point"])
cleanedset
```

Insert answer here  
28.963414634146343

2.5. Correct the following codes so that they do not return errors.  
(Python Equivalents underneath faulty R script)

Original R Code

```
sum(curry$point)
```

Python

```
sum(curry.point)
```

Describe the issue you corrected here.  
Point should be capitalized

Answer

```
sum(curry.Point)
```

Original R Code

```
length(curry$Game
```

Python

```
len(curry.Game
```

Describe the issue you corrected here.  
Missing a Parenthesis

Answer

```
len(curry.Game)
```

Original R Code

```
sum(curry$Point na.rm=TRUE)
```

Python

```
#assume that the parameter na.rm removes all nan from the dataset  
sum(curry.Point na.rm=True)
```

Describe the issue you corrected here.

Add a comma

Answer

```
sum(curry.Point, na.rm=True)
```

---

## Part 3: Government Shutdowns

---

In United States, a government shutdown occurs when Congress fails to pass a bill or a continuing resolution to fund federal government operations and agencies or when the President refuses to sign such a bill or resolution. During a shutdown, some federal employees must take a furlough (temporary leave of employees), and some federal employees are still required to work without pay until the shutdown ends. The duration of a shutdown is random depending, and it prolongs when a conflict between Congress and the President cannot find a compromising solution. `shutdown.csv` includes records of the 10 federal government shutdowns that included furloughs. It includes the year of the shutdown, the sitting President, the number of days the shutdown lasted, the number of furloughed workers and the cost (in millions) of the shutdown.

3.1. Modify the code below by adding in the appropriate information to the blanks to read in the data from `shutdown.csv` and store it as the object `shutdown`.

```
shutdown = pd.read_csv("shutdown.csv")
display(shutdown)
```

3.2. What is the mean number of days a shutdown that includes furloughs has lasted? Modify the inputs for the function below.

```
avgDays = sum(shutdown.Days)/len(shutdown.Days)
```

Insert answer here

8.8 Days

3.3. What is the total cost to Americans for furlough shutdowns, over history?

```
cleanedset = shutdown.values.tolist()
for i in range(len(cleanedset)):
    if math.isnan(cleanedset[i][4]):
        cleanedset[i][4] = 0
shutdown = pd.DataFrame(cleanedset, columns=["Year", "President", "Days", "EmployeesFurloughed", "CostMill"])
cleanedset

totalLoss = sum(shutdown.CostMill)
totalLoss
```

Insert answer here

2715.27

3.4. Use the function `tally()` to create a table of the number of shutdowns under each President. Which Presidents have had the most shutdowns involving furloughs?

```
tally = dict()
for value in shutdown.President.tolist():
    if value in tally:
```

```
tally[value] += 1
else:
    tally[value] = 1
```

Insert answer here

1)Regan  
2)Clinton  
2)Trump  
3)Obama  
3)HWBush  
3)Carter

---

## Part 4: Bison Population Surveys

---

Yellowstone National Park conducts aerial and ground surveys each year of the bison population within the park boundary. They record the number of individuals that are male, female, and juveniles. The records since 2000 are in `bison_yellowstone.csv`.

4.1. Modify the code chunk below to read in your data and store it as `bison`

```
bison = pd.read_csv("bison_yellowstone.csv")
```

4.2. What is the mean male population size since 2000? Insert a code chunk and write the appropriate code to answer the question. Recall, you can enter a code chunk by using the following shortcuts: Windows OS- press Ctrl+Alt+I; for Mac OS Cmd+Alt+I.

```
avgMalePop = sum(bison.Male)/len(bison.Male)
```

Insert answer here.

1378.6666666666667

4.3. Which sex had the highest adult population size? Use the function `max()` (uses the formula template) to answer this question.

```
max(avgMalePop, sum(bison.Female)/len(bison.Male))
```

Insert answer here.

1378.6666666666667

4.4. How many juveniles have been born in Yellowstone since 2000?

```
sum(bison.Juveniles)
```

Insert answer here.

20011

4.5. Researchers are interested in the total population size for each year. Add across vectors within each element and store the new vector as the object `total`. Print `total` to screen to see the total population sizes for each year.

```
population = bison.values.tolist()
for i in range(len(population)):
    population[i].append(sum([population[i][1],population[i][2],population[i][3]]))
bison = pd.DataFrame(population,columns=["Year", "Male", "Female", "Juveniles", "Total"])
```

---

## Part 5: Fish Surveys

---

You are helping Reef Check do fish surveys off the California coast. The survey protocol says the divers should record the species of each fish they see along a transect and the fish's length, in centimeters. Two divers have asked for your help entering their field sheet into an electronic file. They tell you that for ease of underwater recording, they entered the data on their field sheet as: the number of fish (at this size) as a list under each species. Their field sheet is in the file

`FishSurvey_PtLobos_fieldsheet.pdf`.

5.1. Create a spreadsheet (.csv) of the data from field data sheets in the file "FishSurvey\_PtLobos\_fieldsheet.pdf". You must include relevant information on fish species and size, as well as the associated metadata for both transects. You can enter the data in Excel (or other spreadsheet program), but be sure to save your file as a .csv (comma separated value; do not save it as a CSV UTF-8 file, a Macintosh csv or a MS dos csv--you want the regular CSV file called "Comma Seperated Value (.csv)"). (*Note: you cannot have spaces in your vector/variable names.*) Name the file according to the file naming convention, with the suffix "\_fish-data.csv" (So Montey Rey would name his .csv file "ReyM\_fish-data.csv".) Upload the file to your project. You will need to select this file to export along with your .rmd and .rproj files to turn in for this lab.

Now check to see that you file knits by clicking the button 'knit' at the top of this pane. If you have an error in your code, the file will not knit. Knitting to html will appear in the 'Viewer' tab in your "Files" pane if you click the gear tab and select "Preview in Viewer Pane."