

BIOS 611 - HW 3

Ana Dulskiy

Question 1

What are the two ways that commands in a unix-y shell can accept input (other than reading from a file on disk)?

- Command line arguments
- A shell script - executable as bash code, takes user input

Question 2

What about output?

Standard output is printed to the terminal, but can also be redirected to a file using the > or >> operator. You can also pipe the standard output of one command to the standard input of another.

Questions 3 & 4

Create a Dockerfile which includes the standard R environment we've been using but also includes the kernlab R package.

Extend that Dockerfile to include the Nice Editor.

```
FROM rocker/verse
MAINTAINER Ana Dulskiy <adulskiy15@gmail.com>
RUN R -e "install.packages('kernlab')"
RUN apt-get update && apt install ne
```

L: 5 C: 1 100% i-----pvu-T-!-----@A* Dockerfile

Question 5

What does the following script print on the standard output?

```
A_VARIABLE=hello
B_VARIABLE=world
echo $A_VARIABLE B_VARIABLE > some-file
cat some-file
```

The script above returns the output

```
hello B_VARIABLE
```

The code above creates the file, "some-file", which contains A_VARIABLE and B_VARIABLE, which have been assigned "hello" and "world", respectively. The echo command calls the two variables and redirects the standard output to the file "some-file." The cat command allows us to view some-file.

The code doesn't return "hello world" because it needs a \$ before B_VARIABLE.

Question 6

What is an IP address (informally) What about a port number (informally)?

An IP address is essentially the address of the system within the network, and the port number is the address of the service within the system.

Question 7

Describe left, inner, and right joins.

Inner join - when you take rows only in the case where the left table and right table match

Left join - when you want every single row from the left table and where there is a match, you want the data from the right table

Right join - the opposite of left join

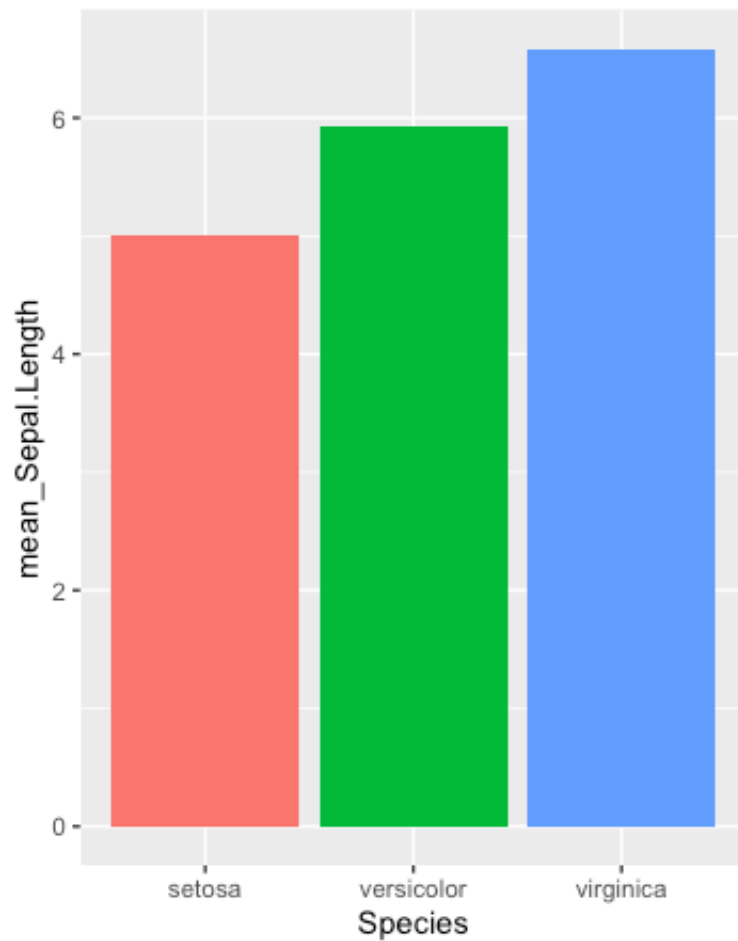
Full join - full pair-wise combination of all the columns

Question 8

R features several built in data sets. One is about flowers and can be accessed via `iris`.

What tidyverse/dplyr code shows the average sepal length for each species? Provide the R code here.

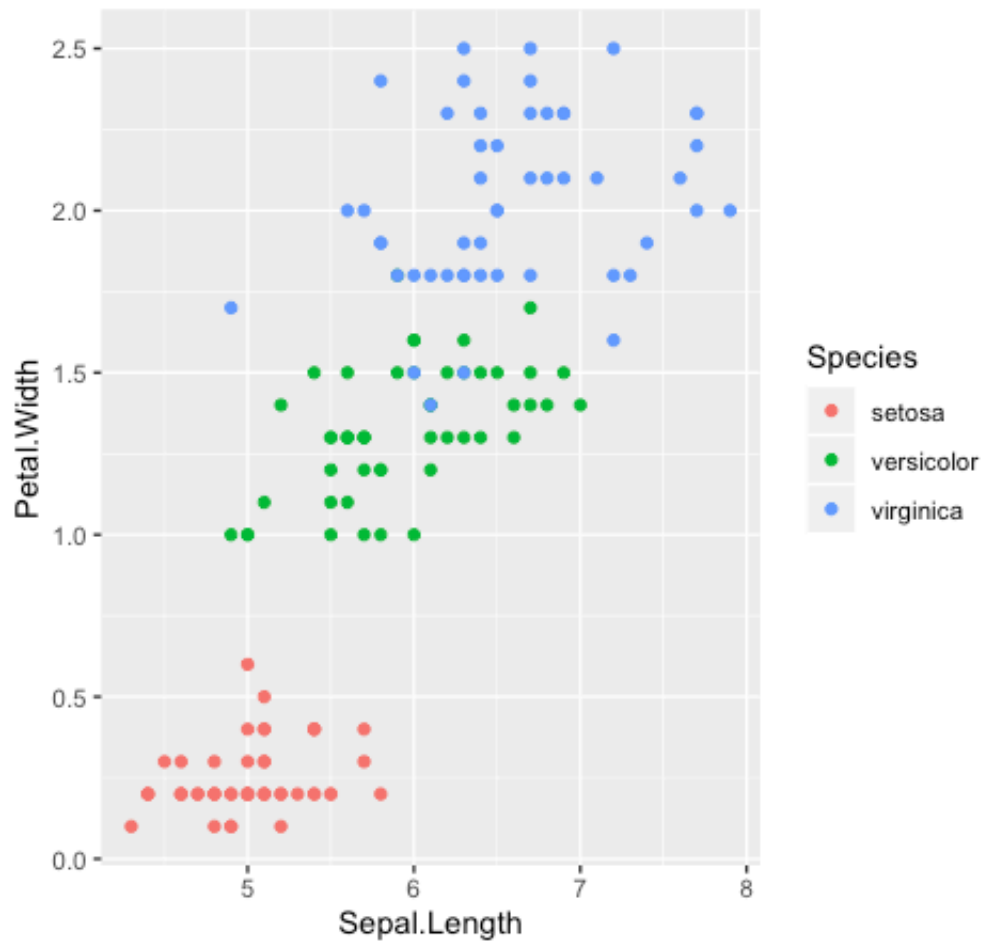
```
iris %>%  
  group_by(Species) %>%  
  summarize(mean_Sepal.Length = mean(Sepal.Length))
```



Question 9

Create a scatter plot of sepal length against petal width. Color code the points by species. Provide the R code which creates the figure.

```
ggplot(iris, aes(x = Sepal.Length, y = Petal.Width, color = Species)) +  
  geom_point()
```



Question 10

Which Species has the smallest sepal length? Recreate the plot in 9 without that species in the data set. Provide R code which filters the data set and produces the figure.

Setosa has the smallest sepal length.

```
iris.vv <- iris %>%  
  filter(Species == "versicolor" | Species == "virginica")  
  
ggplot(iris.vv, aes(x = Sepal.Length, y = Petal.Width, color = Species)) +  
  geom_point()
```

