Lecture 2

Paul Holaway, Abhi Thanvi

June 22nd, 2022

Lecture 1 Review

Before we move on to the new material, we will do a quick review of Lecture 1 content. As we mentioned last time, `R` is basically a fancy calculator that can do many amazing things. However, before we can go onto more complicated topics, let's review a few basics.

Example 1; Assigning and Printing Variables

Remember that when we assign variables, we are saving a numeric value so we can use it later. Also remember, to do this, we do `variable = numeric value` or `variable = expression`. Remember that to run the code cells, click on the green arrow button in the top right-hand corner of the cell.

```
x = 4
fish = 2 + 2
```

To print the variables, remember to type them on their own line.

```
x
```

```
## [1] 4
```

```
fish
```

```
## [1] 4
```

Example 2; Complex Expressions with Variables

We can also use variables in complex expressions to save us the trouble of typing all the numbers out. The examples from last lecture and lab were simple so I will do a more complicated example. Let's say you need to calculate this formula $T = \frac{p}{4r}$. If the numbers are going to be messy, let's use variable assignment to make it easier. Let's say that p and r are know with $p = ln(4)$ and $r = \sqrt[3]{2}$. Remember to hover over the LaTex text to view the expression

```
#Variable Assignment
p = log(4)
r = 2^(1/3)
#Calculation
T = p/(4*r)
T
```

```
## [1] 0.2750756
```

Okay, now we can move onto the next portion of lecture content.

Data Frames and Conditionals

Data Sets and Data Frames

- **Data Set:** A collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer.
- **Data Frame:** A data structure that organizes data into a 2-dimensional table of rows and columns, much like a spreadsheet.

Above are the dictionary definitions, but they are not that intuitive if this is the first time you are being exposed to them. Think of a data set as a collection of data. However, data sets may or may not be well organized (technical terms being structured or unstructured data). Data Frames are as it says in the definition, organized into a 2D table that is easy to read and work with. There is a saying in Statistics and Data Science that 80% of your work is cleaning your data. What this saying means is that about 80% of the work we do it take a disorganized data set and create an organized data frame to work with. While you are not there quite yet, let's begin with first looking at a data set.

Reading A Data Set

We are now going to learn how to read a data set into `RStudio`. It is pretty simple to do once you have done it a few times. Read and follow the instructions below carefully.

1. Go to `Session` at the top bar.
2. Scroll down to `Set Working Directory` and click on `Choose Directory...`.
3. Choose your DPI folder on your computer.

Steps 1-3 will only have to be done if you have closed `RStudio`. Once you set your working directory, it will remain that way until you close `RStudio` where it resets back to the `Desktop`.

4. Download the data set you want. It can either be from Blackboard or another site.
5. Move the data set to your DPI folder on your computer.
6. Click on `Import Dataset`
7. Click on the file type you are trying to import. If you are using…

- a `.csv` or `.tsv` file, then click `From Text (base)...`
- an Excel spreadsheet, then click `From Excel...`

8. Find the file you want to import. All of the data sets for the summer will be uploaded to Blackboard for you to download. It is highly recommended that you keep all of your data sets in the course folder on your computer.
9. Select the file you want to import.

You will only have to follow steps 10-11 if you are importing a `.csv` file. If it is an Excel file, you may skip to step 12.

10. A new window will open up giving you import options, click `Yes` for `Heading` and check the box next to `Strings as factors`.
11. Rename the data set if you wish.
12. Click `Import`

13. Copy and paste the import code in the console into the blank code cell provided. This is necessary so you can convert your labs into PDF format.

If everything goes correctly, you will see the data set open in the top left-hand window and the data set will appear in your local environment. You can view the data set by clicking on it. You most likely will not use any other data types in this course. If you do, ask your instructors for help with importing it into RStudio.

Example 3; Reading a CSV File

Let's follow the instructions above to import a .csv file.

```
hello_csv <- read.csv("~/Classes/DPISu22/Data Sets/hello.csv", stringsAsFactors=TRUE)
```

Excellent, now you have the data imported for you to work with. While .csv files are the standard file type, there are other file types out there. Another common one is an Excel file or .xlsx file.

Example 4; Reading an Excel File

```
library(readxl)
hello_xlsx <- read_excel("Data Sets/hello.xlsx")
```

Notice how here you need a separate package to import Excel data sets. If you do not already have readxl installed, you will need to do so. Refer back to Lecture 1 notes on how to install a package. Other data set types may also require separate packages which is why we are mostly going to give you .csv files to work with. Excel files are common enough where you will need to know how to import them.

Understanding A Data Frame

Now that we have imported the data, let's take a look at it. Notice how we have rows and columns. The hello data set is a survey taken from the UIUC STAT107 students at the beginning of the semester. Each row is an individual observation. So each row is how a student responded to each question. Each column is an observation for a certain attribute for each observation. So a column is how each student responded to a question. You will notice some columns are numeric while some are strings (words). These are the two different types of responses you can have in data. Each one has certain ways that it can be treated. There are cases where they can be treated similarly, while some completely different. We will explore those late on throughout the course. Now let's start playing around.

Tidyverse

Welcome to the meat (or whatever the equivalent is for vegetarians) of the course, the tidyverse. tidyverse is a package that is composed of other packages. The packages in tidyverse are all together because they have been considered some of the most useful and most widely downloaded packages in all of R. So they were incorporated together in one download for convenience. Think of it as a collection of the most useful tools in R in one download. Those packages are…

- ggplot2
- tibble
- tidyr

- readr
- purrr
- dplyr
- stringr
- forcats

Before going onto the examples, install the tidyverse packages. We will be using some, but not all of these. There is unfortunately not enough time to go through everything. Note that we have cheat sheets for dplyr, ggplot2, readr, stringr, and tidyr on Blackboard. These cheat sheets have syntax for all the functions in the package. It may be a bit overwhelming at first to read, but take some time to read through them carefully if you are stuck. Now that we have done that, let's move onto actually doing things.

Summarizing Data

One of the most useful tools in data science is looking at a summary of the data. It will include useful information such as the min, average, and max of numerical data and the frequency for categorical (string) data. This can be done by simply using the summary() function.

Example 5; Summarizing Data

Let's look at the summary of the data. The syntax will be summary(data).

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
hello <- read.csv("~/Classes/DPISu22/Data Sets/hello.csv", stringsAsFactors=TRUE)
```

```
summary(hello)
```

```
##       Name                          Major            Year        Phone
##  Alex   :  2   Information Sciences: 18   Freshman :79   Android: 37
##  Ana    :  2   Statistics          : 16   Junior   :31   iPhone :154
##  Andrew :  2   Information Science : 14   Other    : 3   Other  :  1
##  Carson :  2   Computer Science    :  7   Senior   :27
##  Frank  :  2   Economics           :  6   Sophomore:52
##  Jacob  :  2   Psychology          :  6
##  (Other):180   (Other)             :125
##                      Computer       Straw         Shoe.Size
##  Mac OS X-based computer:101   Min.   :0.000   Min.   : 5.000
##  Windows-based computer : 91   1st Qu.:1.000   1st Qu.: 8.000
##                                Median :1.000   Median : 9.500  
##                                Mean   :1.354   Mean   : 9.435  
##                                3rd Qu.:2.000   3rd Qu.:10.500  
##                                Max.   :2.000   Max.   :44.000  
##                                                                
##       Pets        Hot.Dog                            Streaming 
##  Min.   : 0.0000   No :127   Youtube                       :43  
##  1st Qu.: 0.0000   Yes: 65   Netflix, Youtube              :31  
##  Median : 0.0000             Netflix, Amazon Prime, Youtube:14  
##  Mean   : 0.7344             Netflix, Twitch, Youtube      :14  
##  3rd Qu.: 1.0000             Netflix, HBO Max, Youtube     :11  
##  Max.   :15.0000             Twitch, Youtube               : 8  
##                              (Other)                       :71  
##  Prior.Programming    Season   Statistics.Courses Programming.Courses
##  No : 52          Fall  :78   Min.   : 0.000     Min.   : 0.000     
##  Yes:140          Spring:37   1st Qu.: 1.000     1st Qu.: 0.000     
##                   Summer:56   Median : 1.000     Median : 1.000     
##                   Winter:21   Mean   : 1.651     Mean   : 1.844     
##                               3rd Qu.: 2.000     3rd Qu.: 3.000     
##                               Max.   :12.000     Max.   :14.000     
##                                                                     
##   Study.Hours        Siblings         Sleep            Shoes       
##  Min.   : 0.000   Min.   :0.000   Min.   : 3.000   Min.   : 1.000  
##  1st Qu.: 2.000   1st Qu.:1.000   1st Qu.: 6.000   1st Qu.: 4.000  
##  Median : 4.000   Median :1.000   Median : 7.000   Median : 7.000  
##  Mean   : 4.219   Mean   :1.302   Mean   : 7.105   Mean   : 8.969  
##  3rd Qu.: 5.000   3rd Qu.:2.000   3rd Qu.: 8.000   3rd Qu.:10.000  
##  Max.   :30.000   Max.   :7.000   Max.   :10.000   Max.   :95.000  
##                                                                    
##      Texts           Personality       Zodiac.Sign
##  Min.   :  0.00   Extrovert: 57   Aquarius   :19  
##  1st Qu.:  5.00   Introvert:135   Cancer     :19  
##  Median :  8.00                   Capricorn  :19  
##  Mean   : 16.64                   Leo        :18  
##  3rd Qu.: 20.00                   Sagittarius:18  
##  Max.   :182.00                   Pisces     :17  
##                                   (Other)    :82
```

Ew… that looks a bit messy. However, everything is there. Except, what if we just wanted to look at the summary for a student's amount of time studying, we can definitely make this cleaner. To do that, we will have to learn how to access a specific column.

Accessing Columns

Let's start out with something simple. Let's say you want to look at certain columns in a data set. Using the `hello` data set, let's say you only wanted to look at a students' major and year. To do this, we use the `tidyverse`. Remember to use the `tidyverse`, we have to call the package. Recall this is done using `library(packagename)`, which in this case would be `library(tidyverse)`.

Example 6; Selecting Columns

Now we can do the actual work of getting the specific columns. This is accomplished using the `select()` function. The syntax for this is as follows…

```
dataset %>% select(c("var1","var2",...))
```

Please make sure you remember to have " " around the variable names. The extra code at the end is to print out only the

```
hello %>% select(c("Name","Year"))
```

```
##           Name      Year
## 1     Mathilde    Senior
## 2         Luke Sophomore
## 3       Johnny    Junior
## 4       miller  Freshman
## 5          Tri    Junior
## 6       Dhruva  Freshman
## 7      Jeffrey  Freshman
## 8        Josue    Senior
## 9       Marcel    Junior
## 10      Odalys    Junior
## 11       Derek  Freshman
## 12      Aditya Sophomore
## 13       Hamiz  Freshman
## 14        ziyi  Freshman
## 15      Eugene    Junior
## 16       Elise  Freshman
## 17         Xin  Freshman
## 18  Juan David Sophomore
## 19       Jacob    Junior
## 20        Paul     Other
## 21    Abraheem    Senior
## 22       Jerry  Freshman
## 23          Ye  Freshman
## 24      Yuchen  Freshman
## 25       Rohan Sophomore
## 26   Tiancheng Sophomore
## 27       Aidan  Freshman
## 28       Jimmy  Freshman
## 29      Claire    Junior
## 30     Natalie Sophomore
## 31     Dayanna    Junior
## 32        Riya    Senior
## 33      Lorena Sophomore
## 34       Jakub Sophomore
## 35       Humza  Freshman
## 36         Lin  Freshman
## 37    Wooseong  Freshman
## 38   Shawna Ye  Freshman
## 39      Angela Sophomore
## 40        Josh Sophomore
## 41      Vaishu    Senior
## 42    Jonathan Sophomore
## 43       Kathy Sophomore
## 44     Francis  Freshman
## 45        Rafi    Junior
## 46      Jarred Sophomore
## 47      Carson    Senior
## 48       Frank Sophomore
## 49     Vincent  Freshman
## 50      George  Freshman
## 51        Kyle  Freshman
## 52        Evan  Freshman
## 53       clara    Junior
## 54        Neil Sophomore
## 55         Sam Sophomore
## 56     Jessica    Senior
## 57       Aaron  Freshman
## 58        Deyi Sophomore
## 59        Brad Sophomore
## 60       Dilan  Freshman
## 61      Justin  Freshman
## 62    Degaulle     Other
## 63       Oscar Sophomore
## 64      Muneeb Sophomore
## 65       Dylan    Senior
## 66        Andy Sophomore
## 67        Trey Sophomore
## 68       Emily  Freshman
## 69       Kayla  Freshman
## 70     Zhiheng    Junior
## 71    Victoria  Freshman
## 72        Binh Sophomore
## 73        Nick  Freshman
## 74     Kaiyuan Sophomore
## 75       Vahey    Junior
## 76   Khushalli  Freshman
## 77     Brianna  Freshman
## 78          Ke    Senior
## 79     Jeffrey Sophomore
## 80       Annie  Freshman
## 81      Keaton    Senior
## 82     Martina    Senior
## 83        Dane  Freshman
## 84       Veena    Junior
## 85        Alex  Freshman
## 86         Ram  Freshman
## 87        Veer  Freshman
## 88       Laila    Junior
## 89      Jackie    Junior
## 90       Jacob    Senior
## 91         Min    Senior
## 92         Max    Senior
## 93      Khushi  Freshman
## 94       Riley Sophomore
## 95      Daniel  Freshman
## 96         Jai  Freshman
## 97     Goutham  Freshman
## 98        Nico    Senior
## 99     Vanessa    Senior
## 100    Madison Sophomore
## 101      Bowen Sophomore
## 102      Kayla  Freshman
## 103   Araditta Sophomore
## 104       Coby    Senior
## 105       Noah    Senior
## 106      Azeem  Freshman
## 107      Milan    Junior
## 108     Hojoon Sophomore
## 109      Dhruv  Freshman
## 110   Caroline    Junior
## 111      Subbu  Freshman
## 112     Justin    Senior
## 113       Josh Sophomore
## 114    Trishla  Freshman
## 115      Zihan     Other
## 116    Binkina  Freshman
## 117     Andrew    Senior
## 118    Raleigh  Freshman
## 119        Ana Sophomore
## 120       Nick  Freshman
## 121       Alex    Junior
## 122    Hansika  Freshman
## 123      Tarun  Freshman
## 124       Kate  Freshman
## 125     Pedram  Freshman
## 126      Conor  Freshman
## 127     Keegan    Junior
## 128    Demitri    Junior
## 129   kristian    Junior
## 130         Si  Freshman
## 131      Irene  Freshman
## 132    Valerie    Junior
## 133       Sean  Freshman
## 134    Jinxiao    Senior
## 135    Shriyal  Freshman
## 136        Zoe Sophomore
## 137     Baseet    Senior
## 138    Natasha    Junior
## 139       Arya  Freshman
## 140        Sam Sophomore
## 141    Anthony  Freshman
## 142      Grace  Freshman
## 143  Katherine    Junior
## 144       Hill    Junior
## 145     Kashni  Freshman
## 146     Austin Sophomore
## 147    Xinming  Freshman
## 148     Harish Sophomore
## 149      Nalin  Freshman
## 150      Sabir  Freshman
## 151    Charlie  Freshman
## 152       Gwyn Sophomore
## 153      Xinyi    Junior
## 154     sofiya  Freshman
## 155      Qiuer Sophomore
## 156     Joshua Sophomore
## 157   Sreelaya Sophomore
## 158      Izaak  Freshman
## 159   Chaeyeon  Freshman
## 160      Maria Sophomore
## 161        Ana Sophomore
## 162     Eliana  Freshman
## 163   Jingyuan Sophomore
## 164     Andrew    Senior
## 165      Barry    Junior
## 166      Frank    Senior
## 167     Dakota Sophomore
## 168      Jayha  Freshman
## 169      Sarah Sophomore
## 170     Marcus Sophomore
## 171    Tavarre Sophomore
## 172    Michael  Freshman
## 173     Subash  Freshman
## 174        Jay    Senior
## 175       Joel  Freshman
## 176   Veronica    Junior
## 177     Thomas    Junior
## 178      Yusuk Sophomore
## 179   Rishitaa Sophomore
## 180    Nandika  Freshman
## 181    Chenhao Sophomore
## 182     Ashton    Junior
## 183    Sankalp  Freshman
## 184        Uli    Junior
## 185       Jake  Freshman
## 186      Julia Sophomore
## 187       Joao    Senior
## 188     Carson  Freshman
## 189 Sri Nithya  Freshman
## 190    brandon    Senior
## 191      Yujie  Freshman
## 192     Shiuli Sophomore
```

What this code is saying is, from `dataset` select columns `var1`, `var2`, etc. `%>%` is called the “Pipe Operator”. It tells `RStudio` that you wish to use a function on the data set. Now while this does a nice job at selecting the specific variables, it does not save it as something. If you want to save the specifically selected variables, you will have to use the same assignment procedure as variables.

```r
subset = hello %>% select(c("Name","Major","Year"))
head(subset, 10)
```

```
##        Name                        Major      Year
## 1  Mathilde Community Health & Chemistry    Senior
## 2      Luke                   Stats & CS Sophomore
## 3    Johnny                        ETMAS    Junior
## 4    miller            stat/data science  Freshman
## 5       Tri                      CS+Math    Junior
## 6    Dhruva         Computer Engineering  Freshman
## 7   Jeffrey          Information Science  Freshman
## 8     Josue                     Business    Senior
## 9    Marcel          Information Science    Junior
## 10   Odalys          Information Science    Junior
```

Now if you click on `subset` in your local environment, you will see it only contains each of those three columns. I have printed out the first 10 observations to save pages when converting to a PDF using a function called `head()`. This function will print out the first `n` observations. The code for using the `head()` function is simple… `head(data, n)`.

Example 7; Deselecting Columns

Now let's say you want every column in a data frame but one or two. Let's say the `hello` data set is confidential and you cannot reveal how people respond. This can be done using the same `select()` function, but with a slight change, you simply put `-c("Var1,"var2,...)`. Note the `-` before `c()`. Let's remove the names now.

```
temp = hello %>% select(-c("Name"))
head(temp, 10)
```

```
##                           Major      Year   Phone                Computer Straw
## 1  Community Health & Chemistry    Senior  iPhone  Windows-based computer     1
## 2                    Stats & CS Sophomore Android  Windows-based computer     1
## 3                         ETMAS    Junior  iPhone Mac OS X-based computer     1
## 4             stat/data science  Freshman  iPhone Mac OS X-based computer     1
## 5                       CS+Math    Junior Android  Windows-based computer     2
## 6          Computer Engineering  Freshman  iPhone  Windows-based computer     1
## 7           Information Science  Freshman  iPhone  Windows-based computer     1
## 8                      Business    Senior  iPhone Mac OS X-based computer     1
## 9           Information Science    Junior  iPhone  Windows-based computer     1
## 10          Information Science    Junior  iPhone Mac OS X-based computer     1
##    Shoe.Size Pets Hot.Dog                         Streaming Prior.Programming
## 1        9.0    0     Yes                           Netflix                No
## 2        9.5    0      No                   Twitch, Youtube               Yes
## 3       10.0    2      No                  Netflix, Youtube                No
## 4       10.0    2      No                  Netflix, HBO Max               Yes
## 5       10.0    0     Yes                   Twitch, Youtube               Yes
## 6       10.0    0      No                  Netflix, Youtube               Yes
## 7       10.5    1      No   Hulu, Netflix, HBO Max, Youtube               Yes
## 8       10.0    0      No            Hulu, Netflix, Youtube                No
## 9       11.0    1      No Netflix, HBO Max, Twitch, Youtube               Yes
## 10       9.5    1     Yes Netflix, HBO Max, Twitch, Youtube               Yes
##    Season Statistics.Courses Programming.Courses Study.Hours Siblings Sleep
## 1    Fall                  3                   0         3.0        1   8.0
## 2  Spring                  4                   2         4.0        1   6.5
## 3    Fall                  1                   0         2.0        1   7.0
## 4  Summer                  0                   2         2.0        2   6.0
## 5    Fall                  3                   6         3.0        2   6.0
## 6  Summer                  1                   3         2.5        1   6.0
## 7  Spring                  0                   4         2.0        1   8.0
## 8  Summer                  2                   1         3.0        6   7.0
## 9  Summer                  4                   4         4.0        1   7.0
## 10 Summer                  1                   1         5.0        1   7.0
##    Shoes Texts Personality Zodiac.Sign
## 1     10     7   Introvert Sagittarius
## 2      4    15   Introvert      Gemini
## 3     10     4   Introvert         Leo
## 4      6    12   Extrovert Sagittarius
## 5      1     0   Introvert       Libra
## 6      5    17   Extrovert   Capricorn
## 7      8    50   Introvert      Cancer
## 8     15     8   Introvert       Virgo
## 9      6     6   Introvert      Pisces
## 10     8     6   Extrovert      Gemini
```

Now say we need to remove the students' Phone preferences for some reason as well as their name.

```
temp = hello %>% select(-c("Name","Phone"))
head(temp, 10)
```

```
##                           Major      Year                Computer Straw
## 1  Community Health & Chemistry    Senior  Windows-based computer     1
## 2                    Stats & CS Sophomore  Windows-based computer     1
## 3                         ETMAS    Junior Mac OS X-based computer     1
## 4             stat/data science  Freshman Mac OS X-based computer     1
## 5                       CS+Math    Junior  Windows-based computer     2
## 6          Computer Engineering  Freshman  Windows-based computer     1
## 7           Information Science  Freshman  Windows-based computer     1
## 8                      Business    Senior Mac OS X-based computer     1
## 9           Information Science    Junior  Windows-based computer     1
## 10          Information Science    Junior Mac OS X-based computer     1
##    Shoe.Size Pets Hot.Dog                         Streaming Prior.Programming
## 1        9.0    0     Yes                           Netflix                No
## 2        9.5    0      No                   Twitch, Youtube               Yes
## 3       10.0    2      No                  Netflix, Youtube                No
## 4       10.0    2      No                  Netflix, HBO Max               Yes
## 5       10.0    0     Yes                   Twitch, Youtube               Yes
## 6       10.0    0      No                  Netflix, Youtube               Yes
## 7       10.5    1      No   Hulu, Netflix, HBO Max, Youtube               Yes
## 8       10.0    0      No            Hulu, Netflix, Youtube                No
## 9       11.0    1      No Netflix, HBO Max, Twitch, Youtube               Yes
## 10       9.5    1     Yes Netflix, HBO Max, Twitch, Youtube               Yes
##    Season Statistics.Courses Programming.Courses Study.Hours Siblings Sleep
## 1    Fall                  3                   0         3.0        1   8.0
## 2  Spring                  4                   2         4.0        1   6.5
## 3    Fall                  1                   0         2.0        1   7.0
## 4  Summer                  0                   2         2.0        2   6.0
## 5    Fall                  3                   6         3.0        2   6.0
## 6  Summer                  1                   3         2.5        1   6.0
## 7  Spring                  0                   4         2.0        1   8.0
## 8  Summer                  2                   1         3.0        6   7.0
## 9  Summer                  4                   4         4.0        1   7.0
## 10 Summer                  1                   1         5.0        1   7.0
##    Shoes Texts Personality Zodiac.Sign
## 1     10     7   Introvert Sagittarius
## 2      4    15   Introvert      Gemini
## 3     10     4   Introvert         Leo
## 4      6    12   Extrovert Sagittarius
## 5      1     0   Introvert       Libra
## 6      5    17   Extrovert   Capricorn
## 7      8    50   Introvert      Cancer
## 8     15     8   Introvert       Virgo
## 9      6     6   Introvert      Pisces
## 10     8     6   Extrovert      Gemini
```

Now let's retry looking at the summary for just the amount of time studying.

```
temp = hello %>% select(c("Study.Hours"))
summary(temp)
```

```
##   Study.Hours
##  Min.   : 0.000
##  1st Qu.: 2.000
##  Median : 4.000
##  Mean   : 4.219
##  3rd Qu.: 5.000
##  Max.   :30.000
```

```
#Alternative Way
summary(hello %>% select(c("Study.Hours")))
```

```
##   Study.Hours
##  Min.   : 0.000
##  1st Qu.: 2.000
##  Median : 4.000
##  Mean   : 4.219
##  3rd Qu.: 5.000
##  Max.   :30.000
```

There, much cleaner and easier to read. Plus you do not have to do any searching through a massive chunk of output.

Accessing Rows

What if you wanted to access a specific row in a data set? You may either wish to look at one specific row or a group of rows that fit a certain criteria. Looking at a specific row will be easier by viewing the data set in the viewing panel, so we will not discuss the coding way here. Instead we will focus on looking at rows that fit a certain criteria.

Example 8; Filtering By Name

Let's say you just wanted to look at data for Freshman in the `hello` data set. This can be accomplished using the `filter()` function. The syntax for this is as follows…

```
data %>% filter(Variable == "Condition")
```

Note how here you need to have " " around the condition, but not the variable. This is because we are looking at a categorical (string) variable.

```
temp = hello %>% filter(Year == "Freshman")
head(temp, 10)
```

```
##       Name                Major     Year   Phone                Computer Straw
## 1   miller    stat/data science Freshman  iPhone Mac OS X-based computer     1
## 2   Dhruva Computer Engineering Freshman  iPhone  Windows-based computer     1
## 3  Jeffrey  Information Science Freshman  iPhone  Windows-based computer     1
## 4    Derek            CS + Stat Freshman  iPhone  Windows-based computer     1
## 5    Hamiz                 Math Freshman Android  Windows-based computer     1
## 6     ziyi          mathematics Freshman  iPhone Mac OS X-based computer     2
## 7    Elise           Statistics Freshman  iPhone  Windows-based computer     1
## 8      Xin           psychology Freshman  iPhone Mac OS X-based computer     2
## 9    Jerry     Computer Science Freshman Android  Windows-based computer     1
## 10      Ye           Statistics Freshman  iPhone  Windows-based computer     2
##    Shoe.Size Pets Hot.Dog                         Streaming Prior.Programming
## 1       10.0    2      No                  Netflix, HBO Max               Yes
## 2       10.0    0      No                  Netflix, Youtube               Yes
## 3       10.5    1      No   Hulu, Netflix, HBO Max, Youtube               Yes
## 4       10.5    2     Yes Netflix, HBO Max, Twitch, Youtube               Yes
## 5       10.5    0      No          Netflix, Twitch, Youtube               Yes
## 6        8.5    1      No                           Youtube                No
## 7       12.0    1      No                  Netflix, Youtube                No
## 8        6.0    0      No             Amazon Prime, Youtube               Yes
## 9        9.0    0     Yes          Netflix, Twitch, Youtube               Yes
## 10       9.5    1      No                           Youtube               Yes
##    Season Statistics.Courses Programming.Courses Study.Hours Siblings Sleep
## 1  Summer                  0                   2         2.0        2  6.00
## 2  Summer                  1                   3         2.5        1  6.00
## 3  Spring                  0                   4         2.0        1  8.00
## 4  Spring                  1                   2         2.5        1  8.00
## 5    Fall                  0                   2         4.0        2  7.00
## 6    Fall                  2                   0         8.0        0  8.00
## 7  Winter                  2                   0         3.0        1  6.00
## 8  Spring                  1                   1         3.0        2  7.00
## 9  Winter                  1                   3         3.0        1  7.69
## 10 Winter                  0                   1         5.0        2  9.00
##    Shoes Texts Personality Zodiac.Sign
## 1      6    12   Extrovert Sagittarius
## 2      5    17   Extrovert   Capricorn
## 3      8    50   Introvert      Cancer
## 4      4    10   Introvert      Cancer
## 5      4     4   Introvert Sagittarius
## 6     10    10   Introvert    Aquarius
## 7     10   127   Introvert   Capricorn
## 8      5     6   Introvert    Aquarius
## 9      6     3   Introvert     Scorpio
## 10     5    13   Extrovert   Capricorn
```

Okay, but what if the variable we want to look at is numeric? In that case, it is similar syntax, just you do **NOT** put “ ” around the condition. Let’s look at students who have no pets.

```r
temp = hello %>% filter(Pets == 0)
head(temp, 10)
```

```
##          Name                        Major      Year   Phone
## 1    Mathilde Community Health & Chemistry    Senior  iPhone
## 2        Luke                   Stats & CS Sophomore Android
## 3         Tri                      CS+Math    Junior Android
## 4      Dhruva         Computer Engineering  Freshman  iPhone
## 5       Josue                     Business    Senior  iPhone
## 6      Aditya          Information Science Sophomore Android
## 7       Hamiz                         Math  Freshman Android
## 8         Xin                   psychology  Freshman  iPhone
## 9  Juan David                   Statistics Sophomore  iPhone
## 10      Jacob         Information Sciences    Junior  iPhone
##                    Computer Straw Shoe.Size Pets Hot.Dog
## 1    Windows-based computer     1       9.0    0     Yes
## 2    Windows-based computer     1       9.5    0      No
## 3    Windows-based computer     2      10.0    0     Yes
## 4    Windows-based computer     1      10.0    0      No
## 5   Mac OS X-based computer     1      10.0    0      No
## 6    Windows-based computer     2       8.0    0      No
## 7    Windows-based computer     1      10.5    0      No
## 8   Mac OS X-based computer     2       6.0    0      No
## 9   Mac OS X-based computer     1      11.0    0      No
## 10  Mac OS X-based computer     1      10.0    0     Yes
##                                         Streaming Prior.Programming Season
## 1                                         Netflix                No   Fall
## 2                                 Twitch, Youtube               Yes Spring
## 3                                 Twitch, Youtube               Yes   Fall
## 4                                Netflix, Youtube               Yes Summer
## 5                          Hulu, Netflix, Youtube                No Summer
## 6            Hulu, Netflix, Amazon Prime, Youtube               Yes Spring
## 7                        Netflix, Twitch, Youtube               Yes   Fall
## 8                           Amazon Prime, Youtube               Yes Spring
## 9                                         Youtube               Yes Winter
## 10 Hulu, Netflix, HBO Max, Amazon Prime, Youtube                No Summer
##    Statistics.Courses Programming.Courses Study.Hours Siblings Sleep Shoes
## 1                   3                   0         3.0        1   8.0    10
## 2                   4                   2         4.0        1   6.5     4
## 3                   3                   6         3.0        2   6.0     1
## 4                   1                   3         2.5        1   6.0     5
## 5                   2                   1         3.0        6   7.0    15
## 6                   0                   1         8.0        1   6.0    12
## 7                   0                   2         4.0        2   7.0     4
## 8                   1                   1         3.0        2   7.0     5
## 9                   1                   4         8.0        2   9.0     4
## 10                  2                   0         4.0        1   7.0    15
##    Texts Personality Zodiac.Sign
## 1      7   Introvert Sagittarius
## 2     15   Introvert      Gemini
## 3      0   Introvert       Libra
## 4     17   Extrovert   Capricorn
## 5      8   Introvert       Virgo
## 6      4   Introvert     Scorpio
## 7      4   Introvert Sagittarius
## 8      6   Introvert    Aquarius
## 9      5   Introvert       Libra
## 10     5   Extrovert      Taurus
```

With numeric, we can do a bit more than categorical. Let's say we want to look at students who get a certain amount of sleep, say more than 6 hours on average.

```
temp = hello %>% filter(Sleep > 6)
head(temp, 10)
```

```
##        Name                        Major      Year   Phone
## 1  Mathilde Community Health & Chemistry    Senior  iPhone
## 2      Luke                   Stats & CS Sophomore Android
## 3    Johnny                        ETMAS    Junior  iPhone
## 4   Jeffrey          Information Science  Freshman  iPhone
## 5     Josue                     Business    Senior  iPhone
## 6    Marcel          Information Science    Junior  iPhone
## 7    Odalys          Information Science    Junior  iPhone
## 8     Derek                    CS + Stat  Freshman  iPhone
## 9     Hamiz                         Math  Freshman Android
## 10     ziyi                  mathematics  Freshman  iPhone
##                   Computer Straw Shoe.Size Pets Hot.Dog
## 1   Windows-based computer     1       9.0    0     Yes
## 2   Windows-based computer     1       9.5    0      No
## 3  Mac OS X-based computer     1      10.0    2      No
## 4   Windows-based computer     1      10.5    1      No
## 5  Mac OS X-based computer     1      10.0    0      No
## 6   Windows-based computer     1      11.0    1      No
## 7  Mac OS X-based computer     1       9.5    1     Yes
## 8   Windows-based computer     1      10.5    2     Yes
## 9   Windows-based computer     1      10.5    0      No
## 10 Mac OS X-based computer     2       8.5    1      No
##                             Streaming Prior.Programming Season
## 1                             Netflix                No   Fall
## 2                     Twitch, Youtube               Yes Spring
## 3                    Netflix, Youtube                No   Fall
## 4    Hulu, Netflix, HBO Max, Youtube               Yes Spring
## 5              Hulu, Netflix, Youtube                No Summer
## 6  Netflix, HBO Max, Twitch, Youtube               Yes Summer
## 7  Netflix, HBO Max, Twitch, Youtube               Yes Summer
## 8  Netflix, HBO Max, Twitch, Youtube               Yes Spring
## 9            Netflix, Twitch, Youtube               Yes   Fall
## 10                            Youtube                No   Fall
##    Statistics.Courses Programming.Courses Study.Hours Siblings Sleep Shoes
## 1                   3                   0         3.0        1   8.0    10
## 2                   4                   2         4.0        1   6.5     4
## 3                   1                   0         2.0        1   7.0    10
## 4                   0                   4         2.0        1   8.0     8
## 5                   2                   1         3.0        6   7.0    15
## 6                   4                   4         4.0        1   7.0     6
## 7                   1                   1         5.0        1   7.0     8
## 8                   1                   2         2.5        1   8.0     4
## 9                   0                   2         4.0        2   7.0     4
## 10                  2                   0         8.0        0   8.0    10
##    Texts Personality Zodiac.Sign
## 1      7   Introvert Sagittarius
## 2     15   Introvert      Gemini
## 3      4   Introvert         Leo
## 4     50   Introvert      Cancer
## 5      8   Introvert       Virgo
## 6      6   Introvert      Pisces
## 7      6   Extrovert      Gemini
## 8     10   Introvert      Cancer
## 9      4   Introvert Sagittarius
## 10    10   Introvert    Aquarius
```

Looking now at the data, all students have more than six hours of sleep. You can manipulate the conditions inside `filter()` for different purposes.

- `==`: Is equal to (Categorical or Numeric)
- `!=`: Not equal to (Categorical or Numeric)
- `>`: Greater than (Numeric)
- `<`: Less than (Numeric)
- `>=`: Greater than or equal to (Numeric)
- `<=`: Less than or equal to (Numeric)

Example 9; Combining Conditions

However, what if you want to do multiple conditions at once? This is easy using `%>%`. I will now look at Freshman who have no pets and who get more than six hours of sleep.

```
temp = hello %>% filter(Year == "Freshman") %>% filter(Pets == 0) %>% filter(Sleep > 6)
temp
```

```
##          Name                                    Major     Year   Phone
## 1       Hamiz                                     Math Freshman Android
## 2         Xin                               psychology Freshman  iPhone
## 3       Jerry                         Computer Science Freshman Android
## 4       Jimmy                                    DS+IS Freshman  iPhone
## 5       Humza                              Mathematics Freshman Android
## 6    Wooseong                      Information science Freshman  iPhone
## 7     Francis                     Information Sciences Freshman  iPhone
## 8      George                      Information science Freshman  iPhone
## 9        Kyle                                 Business Freshman  iPhone
## 10      Aaron                         CS + Advertising Freshman Android
## 11      Dilan                     Information Sciences Freshman  iPhone
## 12      Emily                   Statistics and English Freshman  iPhone
## 13  Khushalli      Data Science + Information Sciences Freshman  iPhone
## 14      Annie                      Information Science Freshman  iPhone
## 15     Khushi                                    Stats Freshman  iPhone
## 16      Kayla                     Information Sciences Freshman  iPhone
## 17      Azeem                               CS + Stats Freshman  iPhone
## 18      Subbu                        Political Science Freshman  iPhone
## 19    Trishla                                       IS Freshman Android
## 20       Nick Philosophy and informatics double major. Freshman  iPhone
## 21    Hansika                      Information Systems Freshman  iPhone
## 22      Tarun                                 CS + GIS Freshman  iPhone
## 23       Kate                                Sociology Freshman  iPhone
## 24      Irene                 Economics and Statistics Freshman  iPhone
## 25    Anthony       Information Science + Econometrics Freshman  iPhone
## 26      Grace                     information sciences Freshman  iPhone
## 27     Kashni                               psychology Freshman  iPhone
## 28    Xinming                         Computer science Freshman  iPhone
## 29      Nalin                       Finance (Business) Freshman Android
## 30    Michael                                     STAT Freshman  iPhone
## 31     Subash                      Information Science Freshman  iPhone
## 32       Joel                               Psychology Freshman  iPhone
## 33    Nandika                     Computer Engineering Freshman  iPhone
## 34    Sankalp                        CS and Statistics Freshman Android
## 35       Jake                                     MACS Freshman Android
## 36 Sri Nithya                     Information Sciences Freshman  iPhone
## 37      Yujie                                      MCB Freshman  iPhone
##                   Computer Straw Shoe.Size Pets Hot.Dog
## 1   Windows-based computer     1      10.5    0      No
## 2  Mac OS X-based computer     2       6.0    0      No
## 3   Windows-based computer     1       9.0    0     Yes
## 4   Windows-based computer     1      10.5    0      No
## 5   Windows-based computer     1       8.5    0      No
## 6   Windows-based computer     1       9.5    0      No
## 7   Windows-based computer     2      10.5    0      No
## 8  Mac OS X-based computer     2      11.0    0      No
## 9  Mac OS X-based computer     2      10.5    0     Yes
## 10 Mac OS X-based computer     2      11.0    0      No
## 11 Mac OS X-based computer     1      13.0    0      No
## 12 Mac OS X-based computer     2       8.0    0      No
## 13 Mac OS X-based computer     2       9.0    0      No
## 14 Mac OS X-based computer     1       8.0    0     Yes
## 15 Mac OS X-based computer     1       6.0    0      No
## 16  Windows-based computer     2      10.5    0      No
## 17 Mac OS X-based computer     1       9.0    0      No
## 18  Windows-based computer     2      11.0    0      No
## 19 Mac OS X-based computer     1       7.0    0      No
## 20  Windows-based computer     1       9.5    0     Yes
## 21  Windows-based computer     2      11.0    0     Yes
## 22 Mac OS X-based computer     2      12.5    0      No
## 23 Mac OS X-based computer     1       6.0    0     Yes
## 24 Mac OS X-based computer     1      10.0    0     Yes
## 25 Mac OS X-based computer     1       9.0    0      No
## 26 Mac OS X-based computer     2       5.5    0      No
## 27 Mac OS X-based computer     1       8.5    0      No
## 28 Mac OS X-based computer     1       9.0    0      No
## 29 Mac OS X-based computer     1      10.0    0     Yes
## 30 Mac OS X-based computer     1       8.0    0      No
## 31  Windows-based computer     1      10.5    0     Yes
## 32  Windows-based computer     2      10.5    0     Yes
## 33 Mac OS X-based computer     0       8.5    0     Yes
## 34 Mac OS X-based computer     2       7.0    0      No
## 35  Windows-based computer     1      10.5    0     Yes
## 36  Windows-based computer     1       5.0    0      No
## 37  Windows-based computer     2       9.0    0      No
##                                        Streaming Prior.Programming Season
## 1                       Netflix, Twitch, Youtube               Yes   Fall
## 2                          Amazon Prime, Youtube               Yes Spring
## 3                       Netflix, Twitch, Youtube               Yes Winter
## 4                               Netflix, Youtube               Yes   Fall
## 5         Netflix, Amazon Prime, Twitch, Youtube               Yes Summer
## 6                                        Youtube               Yes   Fall
## 7                               Netflix, Youtube               Yes Spring
## 8                               Netflix, Youtube               Yes Winter
## 9                      Netflix, HBO Max, Youtube               Yes Summer
## 10                                       Youtube               Yes Spring
## 11                        Hulu, Netflix, Youtube               Yes   Fall
## 12                     Netflix, HBO Max, Youtube                No   Fall
## 13                                       Youtube               Yes Summer
## 14                                       Youtube                No   Fall
## 15                              Netflix, Youtube               Yes Summer
## 16                              Netflix, Youtube               Yes Spring
## 17                                       Youtube               Yes   Fall
## 18                                       Youtube               Yes Summer
## 19                         Amazon Prime, Youtube               Yes Winter
## 20                               Twitch, Youtube               Yes Winter
## 21                                 Hulu, Netflix               Yes   Fall
## 22                Netflix, Amazon Prime, Youtube                No Spring
## 23                        Hulu, Netflix, HBO Max                No   Fall
## 24                     Netflix, HBO Max, Youtube                No Spring
## 25                              Netflix, Youtube               Yes Winter
## 26                              Netflix, Youtube               Yes   Fall
## 27                     Netflix, HBO Max, Youtube                No Summer
## 28                                       Youtube               Yes Spring
## 29                                       Youtube               Yes   Fall
## 30                                       Youtube               Yes   Fall
## 31                              Netflix, Youtube               Yes Winter
## 32                Netflix, Amazon Prime, Youtube                No Summer
## 33                Netflix, Amazon Prime, Youtube               Yes   Fall
## 34 Hulu, Netflix, Amazon Prime, Twitch, Youtube               Yes   Fall
## 35             Netflix, HBO Max, Twitch, Youtube               Yes Winter
## 36       Netflix, HBO Max, Amazon Prime, Youtube               Yes   Fall
## 37                                       Youtube               Yes   Fall
##    Statistics.Courses Programming.Courses Study.Hours Siblings Sleep Shoes
## 1                   0                   2         4.0        2  7.00     4
## 2                   1                   1         3.0        2  7.00     5
## 3                   1                   3         3.0        1  7.69     6
## 4                   2                   3         1.0        1  7.00     2
## 5                   1                   1         5.0        1  7.00     5
## 6                   2                   3         4.0        1  8.00     2
## 7                   0                   1         1.0        1  8.00     4
## 8                   0                   1         1.0        3  7.00    10
## 9                   2                   1         1.0        2  7.00     5
## 10                  1                   2         2.0        1  7.00    10
## 11                  2                   5         1.0        1  7.00     3
## 12                  1                   0         3.0        1  8.00    40
## 13                  1                   3         3.0        1  8.00    10
## 14                  1                   0         8.0        0  7.50     6
## 15                  2                   1         6.0        1  7.00    10
## 16                  0                   2         5.0        2  7.00     5
## 17                  1                   1         3.0        1  7.00     6
## 18                  0                   1         3.0        1  8.00     8
## 19                  2                   1         3.0        1  8.00    27
## 20                  1                   4         3.0        0  7.00     4
## 21                  1                   3         3.0        1  7.00     8
## 22                  0                   3         3.0        1  8.00     5
## 23                  1                   0         3.0        1  7.00    12
## 24                  2                   0         3.0        1  8.00    10
## 25                  0                   2         6.0        0  8.00     2
## 26                  1                   0         4.0        1  8.00     8
## 27                  2                   0         3.0        2  8.00    15
## 28                  0                   1         4.0        1  8.00     8
## 29                  0                   2         3.0        1  8.00     3
## 30                  1                   1         2.0        0  8.00     5
## 31                  0                   1         2.0        1  7.50     2
## 32                  1                   0         1.5        2  6.50     4
## 33                  0                   3         3.0        1  7.00    10
## 34                  1                   1         5.0        1  7.00     4
## 35                  0                   0         2.0        1  7.00     6
## 36                  0                   0         5.0        1  7.00    20
## 37                  2                   0         6.0        4  7.00     5
##    Texts Personality Zodiac.Sign
## 1      4   Introvert Sagittarius
## 2      6   Introvert    Aquarius
## 3      3   Introvert     Scorpio
## 4     15   Introvert         Leo
## 5      2   Introvert       Aries
## 6      3   Introvert       Aries
## 7     13   Introvert       Aries
## 8      5   Introvert      Gemini
## 9     75   Introvert       Libra
## 10     3   Introvert         Leo
## 11    30   Introvert Sagittarius
## 12    19   Introvert   Capricorn
## 13     8   Extrovert    Aquarius
## 14    10   Extrovert      Cancer
## 15    15   Introvert Sagittarius
## 16     6   Introvert       Aries
## 17    33   Introvert      Cancer
## 18     7   Introvert Sagittarius
## 19    10   Extrovert      Taurus
## 20     2   Introvert      Cancer
## 21     8   Extrovert Sagittarius
## 22    24   Introvert      Taurus
## 23     8   Extrovert      Pisces
## 24    20   Introvert      Pisces
## 25     7   Introvert   Capricorn
## 26     6   Introvert      Taurus
## 27    10   Extrovert      Taurus
## 28     6   Introvert      Pisces
## 29     7   Introvert      Taurus
## 30    10   Introvert      Cancer
## 31     4   Introvert       Aries
## 32     5   Extrovert      Gemini
## 33    30   Introvert    Aquarius
## 34     5   Introvert     Scorpio
## 35    40   Extrovert      Taurus
## 36    45   Extrovert       Virgo
## 37     6   Introvert     Scorpio
```

Looks like 37 people, or about 19.27% of the people in the data fit these criterion.

End of Lecture 2 Notes