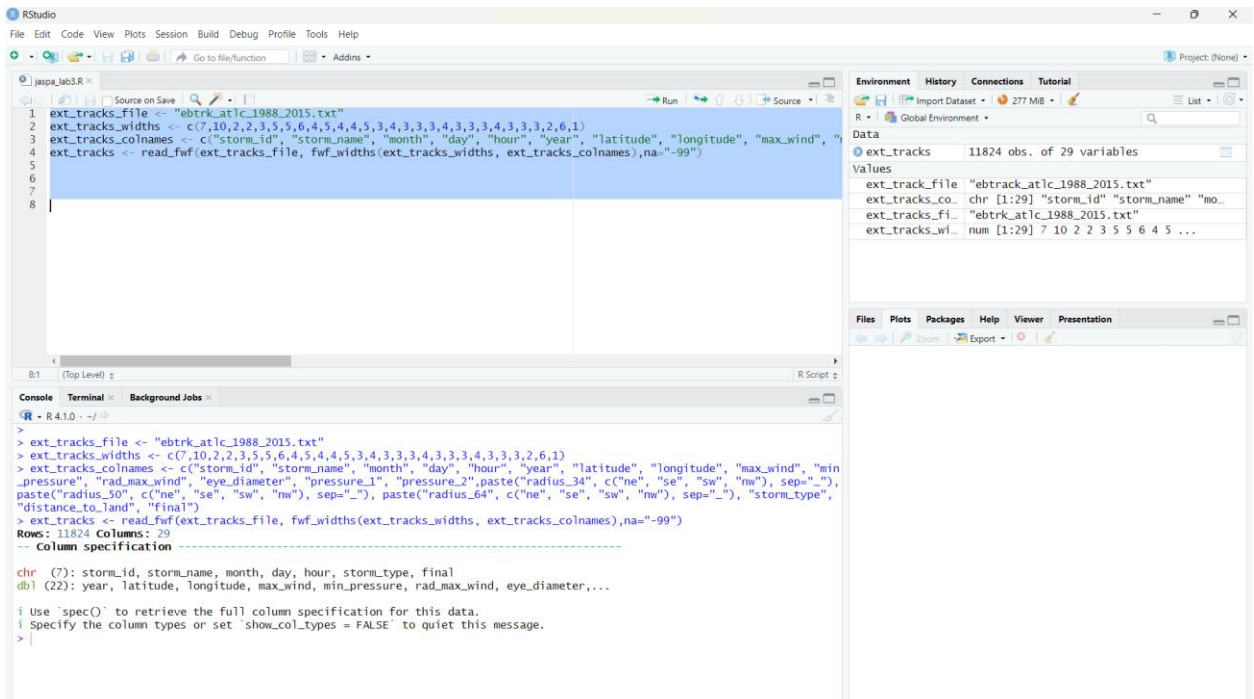


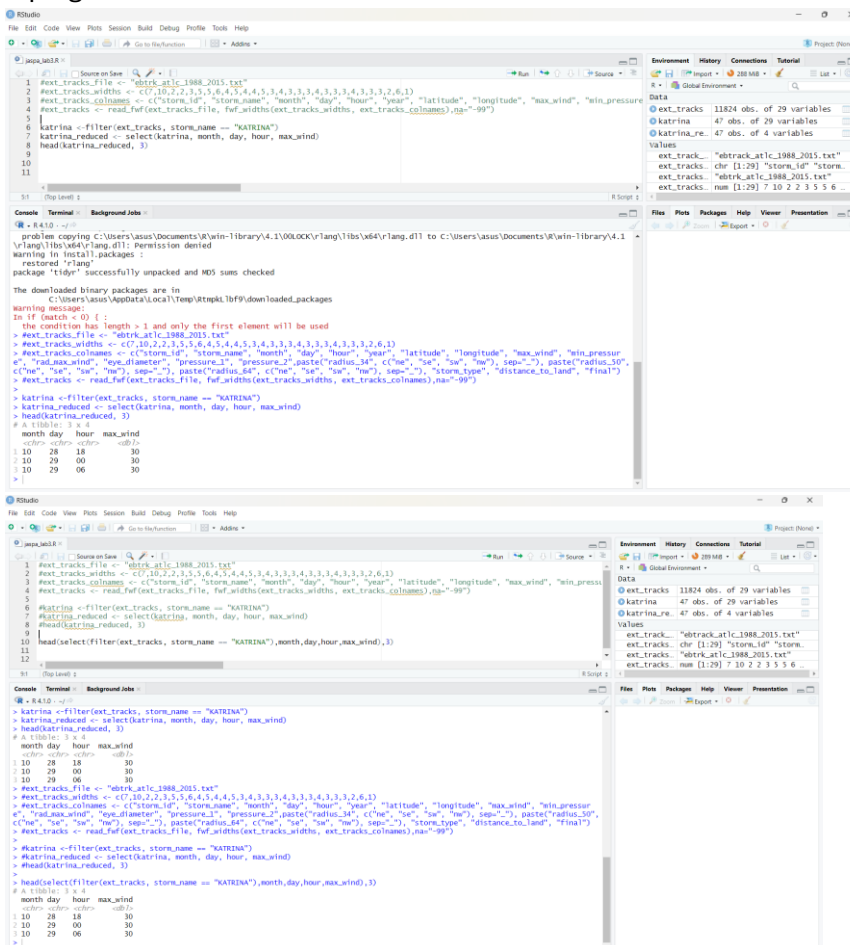
## 1.) Code



```
1 ext_tracks_file <- "ebtrk_atlc_1988-2015.txt"
2 ext_tracks_widths <- c(7,10,2,2,3,5,5,6,4,5,4,4,5,3,4,3,3,4,3,3,4,3,3,2,6,1)
3 ext_tracks_colnames <- c("storm_id", "storm_name", "month", "day", "hour", "year", "latitude", "longitude", "max_wind", "min_pressure", "rad_max_wind", "eye_diameter", "pressure_1", "pressure_2", paste("radius_34", c("ne", "se", "sw", "nw"), sep="_"), paste("radius_50", c("ne", "se", "sw", "nw"), sep="_"), paste("radius_64", c("ne", "se", "sw", "nw"), sep="_"), "storm_type", "distance_to_land", "final")
4 ext_tracks <- read_fwf(ext_tracks_file, fwf_widths(ext_tracks_widths, ext_tracks_colnames), na="-99")
5
6
7
8
```

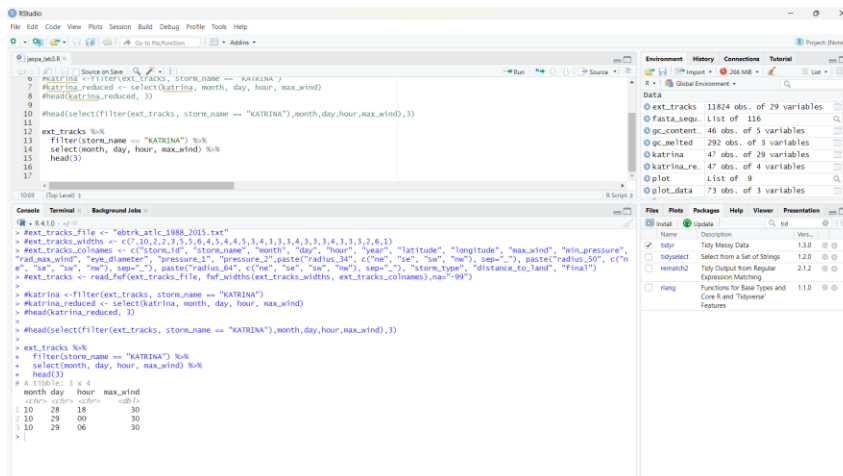
Rows: 11824 Columns: 29  
-- Column specification -----  
chr (7): storm\_id, storm\_name, month, day, hour, storm\_type, final  
dbl (22): year, latitude, longitude, max\_wind, min\_pressure, rad\_max\_wind, eye\_diameter, ...  
i Use 'spec()' to retrieve the full column specification for this data.  
i Specify the column types or set 'show\_col\_types = FALSE' to quiet this message.

## 2.) Piping

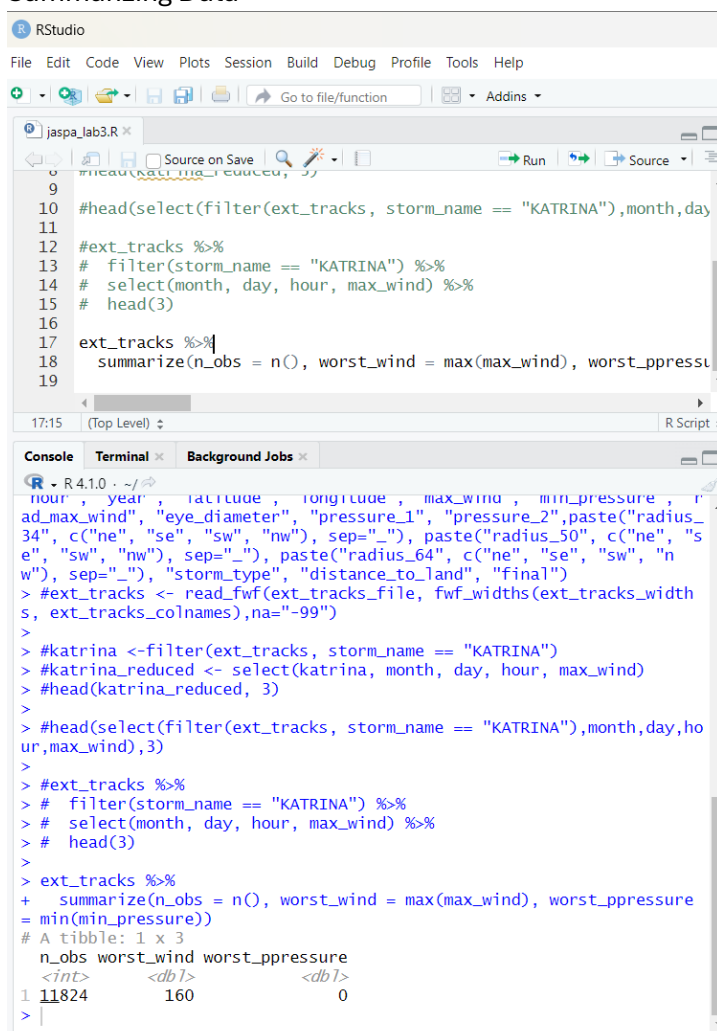


```
1 ext_tracks_file <- "ebtrk_atlc_1988-2015.txt"
2 ext_tracks_widths <- c(7,10,2,2,3,5,5,6,4,5,4,4,5,3,4,3,3,4,3,3,4,3,3,2,6,1)
3 ext_tracks_colnames <- c("storm_id", "storm_name", "month", "day", "hour", "year", "latitude", "longitude", "max_wind", "min_pressure", "rad_max_wind", "eye_diameter", "pressure_1", "pressure_2", paste("radius_34", c("ne", "se", "sw", "nw"), sep="_"), paste("radius_50", c("ne", "se", "sw", "nw"), sep="_"), paste("radius_64", c("ne", "se", "sw", "nw"), sep="_"), "storm_type", "distance_to_land", "final")
4 ext_tracks <- read_fwf(ext_tracks_file, fwf_widths(ext_tracks_widths, ext_tracks_colnames), na="-99")
5
6 # Katrina
7 katrina <- filter(ext_tracks, storm_name == "KATRINA")
8 katrina_reduced <- select(katrina, month, day, hour, max_wind)
9 head(katrina_reduced, 3)
10
11 # A tibble: 3 x 4
12   month day hour max_wind
13   <dbl> <dbl> <dbl> <dbl>
14 1 10 28 18 30
15 2 10 29 00 30
16 3 10 29 06 30
```

```
1 ext_tracks_file <- "ebtrk_atlc_1988-2015.txt"
2 ext_tracks_widths <- c(7,10,2,2,3,5,5,6,4,5,4,4,5,3,4,3,3,4,3,3,4,3,3,2,6,1)
3 ext_tracks_colnames <- c("storm_id", "storm_name", "month", "day", "hour", "year", "latitude", "longitude", "max_wind", "min_pressure", "rad_max_wind", "eye_diameter", "pressure_1", "pressure_2", paste("radius_34", c("ne", "se", "sw", "nw"), sep="_"), paste("radius_50", c("ne", "se", "sw", "nw"), sep="_"), paste("radius_64", c("ne", "se", "sw", "nw"), sep="_"), "storm_type", "distance_to_land", "final")
4 ext_tracks <- read_fwf(ext_tracks_file, fwf_widths(ext_tracks_widths, ext_tracks_colnames), na="-99")
5
6 # Katrina
7 katrina <- filter(ext_tracks, storm_name == "KATRINA")
8 katrina_reduced <- select(katrina, month, day, hour, max_wind)
9 head(katrina_reduced, 3)
10
11 # A tibble: 3 x 4
12   month day hour max_wind
13   <dbl> <dbl> <dbl> <dbl>
14 1 10 28 18 30
15 2 10 29 00 30
16 3 10 29 06 30
```

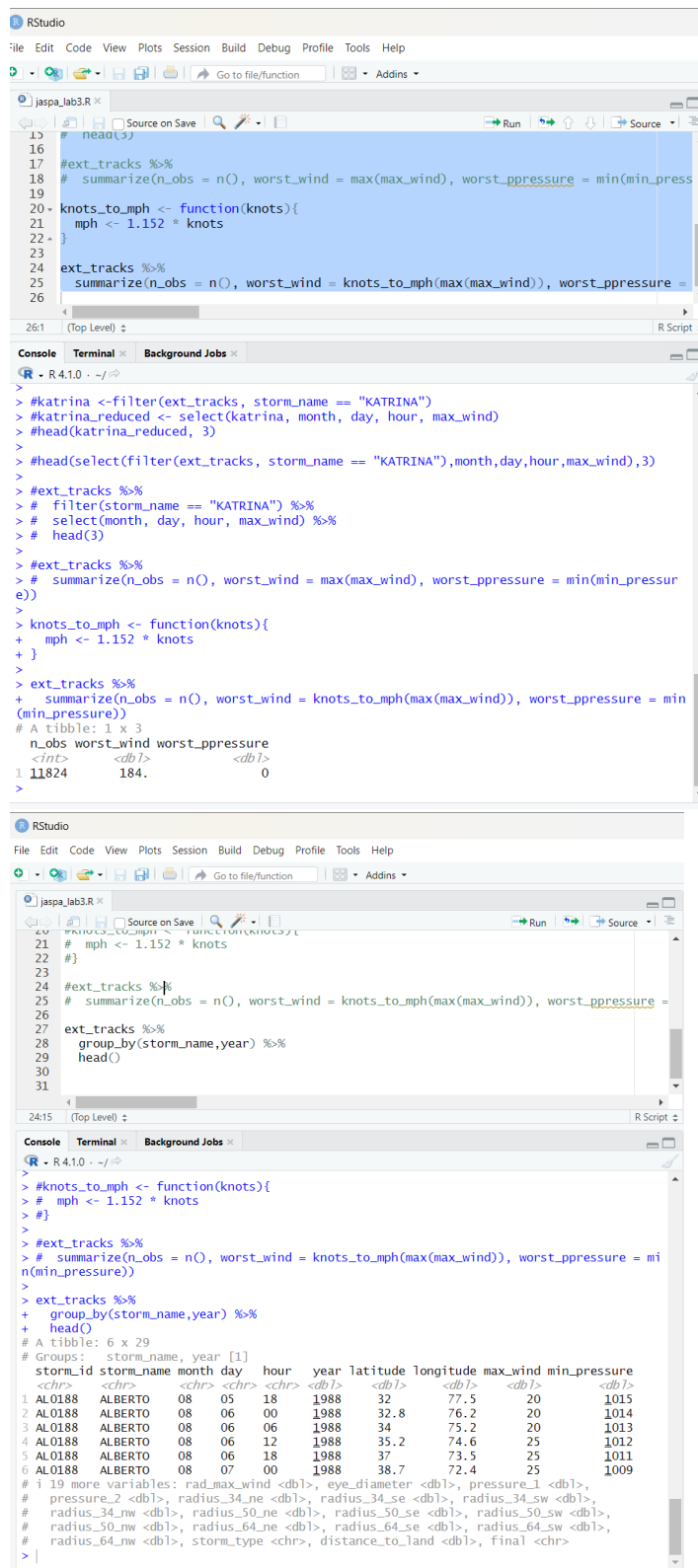


### 3.) Summarizing Data



## Adrian Joel Jaspa

### CMSC 124 - LAB



The image displays two screenshots of an RStudio interface, showing R code being executed and the resulting console output.

**Top Screenshot:**

```
# jaspa_lab3.R
12 # head(3)
16
17 #ext_tracks %>%
18 # summarize(n_obs = n(), worst_wind = max(max_wind), worst_ppressure = min(min_press
19
20 knots_to_mph <- function(knots){
21   mph <- 1.152 * knots
22 }
23
24 ext_tracks %>%
25 summarize(n_obs = n(), worst_wind = knots_to_mph(max(max_wind)), worst_ppressure =
26
```

Console output:

```
>
> #katrina <- filter(ext_tracks, storm_name == "KATRINA")
> #katrina_reduced <- select(katrina, month, day, hour, max_wind)
> #head(katrina_reduced, 3)
>
> #head(select(filter(ext_tracks, storm_name == "KATRINA"), month, day, hour, max_wind), 3)
>
> #ext_tracks %>%
> # filter(storm_name == "KATRINA") %>%
> # select(month, day, hour, max_wind) %>%
> # head(3)
>
> #ext_tracks %>%
> # summarize(n_obs = n(), worst_wind = max(max_wind), worst_ppressure = min(min_pressur
> e))
>
> knots_to_mph <- function(knots){
+   mph <- 1.152 * knots
+ }
>
> ext_tracks %>%
+ summarize(n_obs = n(), worst_wind = knots_to_mph(max(max_wind)), worst_ppressure = min
+ (min_pressure))
# A tibble: 1 x 3
  n_obs worst_wind worst_ppressure
<int>   <dbl>         <dbl>
1 11824      184.           0
```

**Bottom Screenshot:**

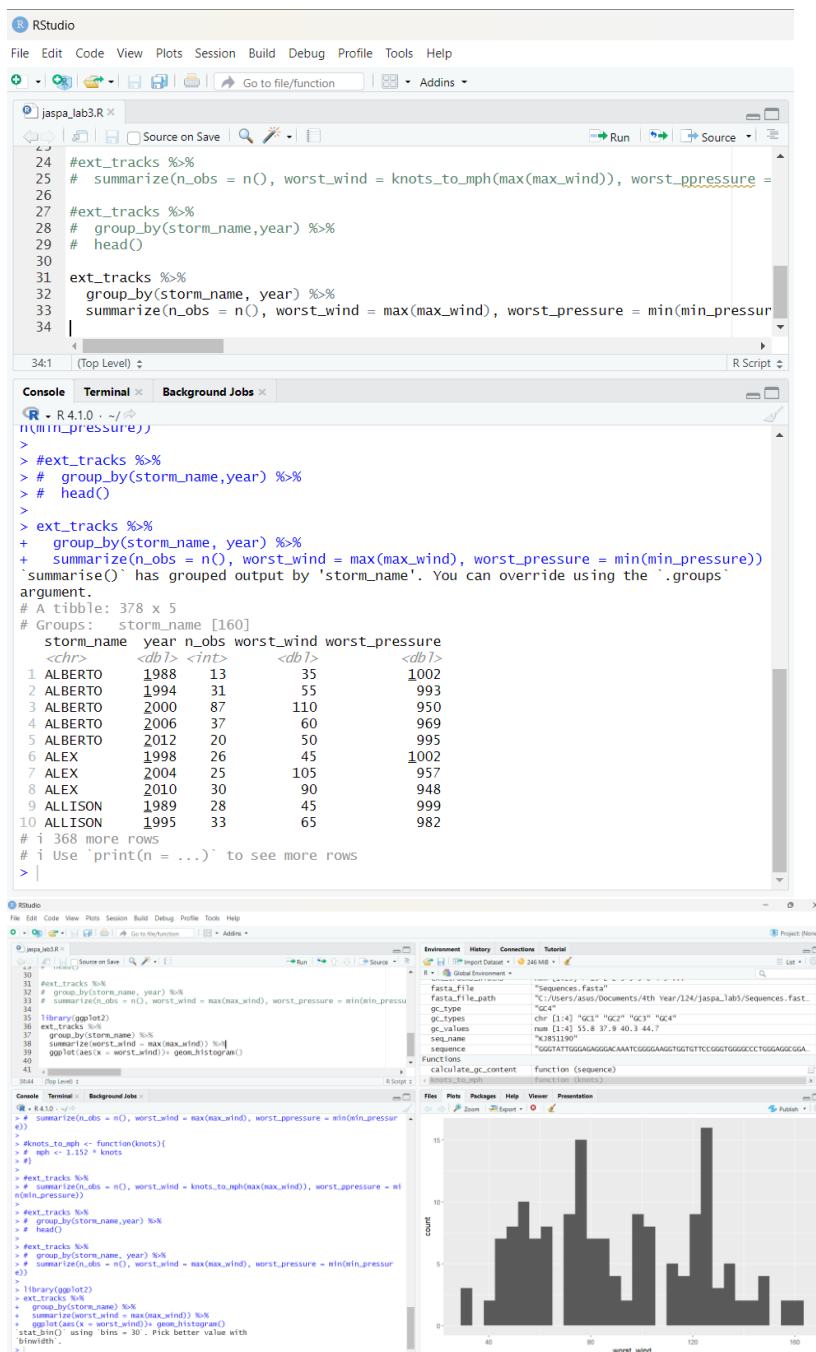
```
20 #knots_to_mph <- function(knots){
21   mph <- 1.152 * knots
22 }
23
24 #ext_tracks %>%
25 # summarize(n_obs = n(), worst_wind = knots_to_mph(max(max_wind)), worst_ppressure =
26
27 ext_tracks %>%
28   group_by(storm_name, year) %>%
29   head()
30
31
```

Console output:

```
>
> #knots_to_mph <- function(knots){
> #   mph <- 1.152 * knots
> #}
>
> #ext_tracks %>%
> # summarize(n_obs = n(), worst_wind = knots_to_mph(max(max_wind)), worst_ppressure = mi
> n(min_pressure))
>
> ext_tracks %>%
+   group_by(storm_name, year) %>%
+   head()
# A tibble: 6 x 29
# Groups:   storm_name, year [1]
  storm_id storm_name month day hour year latitude longitude max_wind min_pressure
<chr>      <chr>      <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
1 AL0188 ALBERTO 08 05 18 1988 32 77.5 20 1015
2 AL0188 ALBERTO 08 06 00 1988 32.8 76.2 20 1014
3 AL0188 ALBERTO 08 06 06 1988 34 75.2 20 1013
4 AL0188 ALBERTO 08 06 12 1988 35.2 74.6 25 1012
5 AL0188 ALBERTO 08 06 18 1988 37 73.5 25 1011
6 AL0188 ALBERTO 08 07 00 1988 38.7 72.4 25 1009
# i 19 more variables: rad_max_wind <dbl>, eye_diameter <dbl>, pressure_1 <dbl>,
# pressure_2 <dbl>, radius_34_ne <dbl>, radius_34_se <dbl>, radius_34_sw <dbl>,
# radius_34_nw <dbl>, radius_50_ne <dbl>, radius_50_se <dbl>, radius_50_sw <dbl>,
# radius_50_nw <dbl>, radius_64_ne <dbl>, radius_64_se <dbl>, radius_64_sw <dbl>,
# radius_64_nw <dbl>, storm_type <chr>, distance_to_land <dbl>, final <chr>
> |
```

## Adrian Joel Jaspa

### CMSC 124 - LAB



#### 4.) Selecting and filtering data

Adrian Joel Jaspa  
CMSC 124 - LAB

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

jaspia_lab3.R
Source on Save Run Source
32 # group_by(storm_name, year) %>%
33 # summarize(n_obs = n(), worst_wind = max(max_wind), worst_pressure = min(min_pressure)) %>%
34
35 #library(ggplot2)
36 #ext_tracks %>%
37 # group_by(storm_name) %>%
38 # summarize(worst_wind = max(max_wind)) %>%
39 # ggplot(aes(x = worst_wind))+ geom_histogram()
40
41 ext_tracks %>%
42 select(storm_name, month, day, hour, year, latitude, longitude, max_wind)
43
41:15 (Top Level) R Script

Console Terminal Background Jobs
R - R4.1.0 - ~/r
e))
>
> #library(ggplot2)
> #ext_tracks %>%
> # group_by(storm_name) %>%
> # summarize(worst_wind = max(max_wind)) %>%
> # ggplot(aes(x = worst_wind))+ geom_histogram()
>
> #ext_tracks %>%
> # select(storm_name, month, day, hour, year, latitude, longitude, max_wind)
# A tibble: 11,824 x 8
  storm_name month day hour year
  <chr> <chr> <chr> <chr> <dbl>
1 ALBERTO 08 05 18 1988
2 ALBERTO 08 06 00 1988
3 ALBERTO 08 06 06 1988
4 ALBERTO 08 06 12 1988
5 ALBERTO 08 06 18 1988
6 ALBERTO 08 07 00 1988
7 ALBERTO 08 07 06 1988
8 ALBERTO 08 07 12 1988
9 ALBERTO 08 07 18 1988
10 ALBERTO 08 08 00 1988
# i 11,814 more rows
# i 3 more variables: latitude <dbl>,
# longitude <dbl>, max_wind <dbl>
# i Use 'print(n = ...)' to see more rows
> |

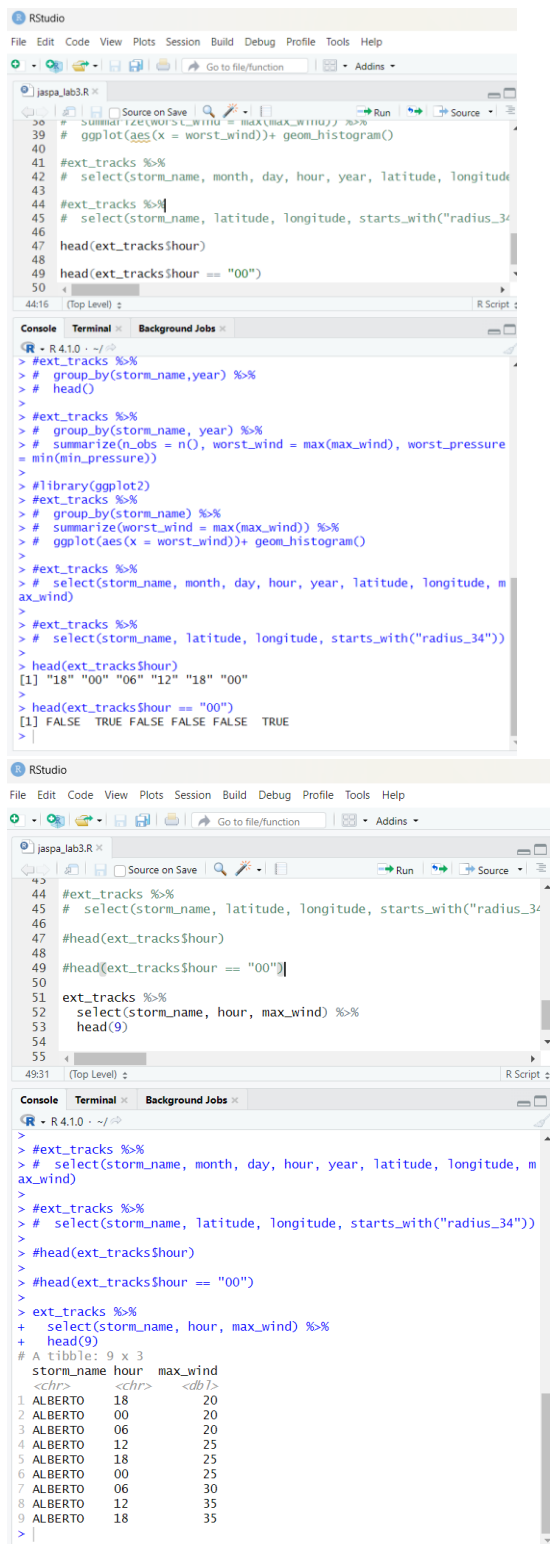
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

jaspia_lab3.R
Source on Save Run Source
35 brary(ggplot2)
36 t_tracks %>%
37 group_by(storm_name) %>%
38 summarize(worst_wind = max(max_wind)) %>%
39 ggplot(aes(x = worst_wind))+ geom_histogram()
40
41
42 select(storm_name, month, day, hour, year, latitude, longitude,
43
44 t_tracks %>%
45 select(storm_name, latitude, longitude, starts_with("radius_34"))
46
46:1 (Top Level) R Script

Console Terminal Background Jobs
R - R4.1.0 - ~/r
> #ext_tracks %>%
> # group_by(storm_name) %>%
> # summarize(worst_wind = max(max_wind)) %>%
> # ggplot(aes(x = worst_wind))+ geom_histogram()
>
> #ext_tracks %>%
> # select(storm_name, month, day, hour, year, latitude, longitude,
max_wind)
>
> #ext_tracks %>%
> # select(storm_name, latitude, longitude, starts_with("radius_34"))
# A tibble: 11,824 x 7
  storm_name latitude longitude radius_34_ne radius_34_se
  <chr> <dbl> <dbl> <dbl> <dbl>
1 ALBERTO 32 77.5 0 0
2 ALBERTO 32.8 76.2 0 0
3 ALBERTO 34 75.2 0 0
4 ALBERTO 35.2 74.6 0 0
5 ALBERTO 37 73.5 0 0
6 ALBERTO 38.7 72.4 0 0
7 ALBERTO 40 70.8 0 0
8 ALBERTO 41.5 69 100 100
9 ALBERTO 43 67.5 100 100
10 ALBERTO 45 65.5 NA NA
# i 11,814 more rows
# i 2 more variables: radius_34_sw <dbl>, radius_34_nw <dbl>
# i Use 'print(n = ...)' to see more rows
> |
```

# Adrian Joel Jaspa

## CMSC 124 - LAB



The image displays two screenshots of the RStudio environment, showing the progression of an R script execution.

**Top Screenshot:**

- Code Editor:** Lines 38-50 of `jaspa_lab3.R` are visible. The code includes a `summarize` call, a `ggplot` call with `geom_histogram`, and a `select` call to filter data by `starts_with("radius_34")`. It also includes `head` calls to inspect the data.
- Console:** Shows the execution of the code. The output of `head(ext_tracks$hour)` is `[1] "18" "00" "06" "12" "18" "00"`. The output of `head(ext_tracks$hour == "00")` is `[1] FALSE TRUE FALSE FALSE FALSE TRUE`.

**Bottom Screenshot:**

- Code Editor:** Lines 44-53 of `jaspa_lab3.R` are visible. The code includes a `select` call, a `head` call, and a `select` call to filter data by `starts_with("radius_34")`. It also includes a `head` call to inspect the data.
- Console:** Shows the execution of the code. The output of `head(ext_tracks$hour)` is `[1] "18" "00" "06" "12" "18" "00"`. The output of `head(ext_tracks$hour == "00")` is `[1] FALSE TRUE FALSE FALSE FALSE TRUE`. The output of `select(storm_name, hour, max_wind)` is a tibble with 9 rows and 3 columns: `storm_name`, `hour`, and `max_wind`. The data is as follows:

storm_name	hour	max_wind
ALBERTO	18	20
ALBERTO	00	20
ALBERTO	06	20
ALBERTO	12	25
ALBERTO	18	25
ALBERTO	00	25
ALBERTO	06	30
ALBERTO	12	35
ALBERTO	18	35

# Adrian Joel Jaspas

## CMSC 124 - LAB

The image displays three sequential screenshots of the RStudio interface, showing the progression of an R script and its execution results in the console.

**Top Left Screenshot:** The script file `jaspas_lab3.R` contains the following code:

```
#head(ext_tracks$hour) %>%
# select(storm_name, hour, max_wind) %>%
# head(9)

ext_tracks %>%
  select(storm_name, hour, max_wind) %>%
  filter(hour == "00") %>%
  head(3)
```

The console shows the execution of `ggplot(aes(x = worst_wind)) + geom_histogram()` and the resulting tibble:

```
# A tibble: 3 x 3
  storm_name hour max_wind
  <chr>      <chr>    <dbl>
1 ALBERTO   00         20
2 ALBERTO   00         25
3 ALBERTO   00         35
```

**Top Right Screenshot:** The script file `jaspas_lab3.R` contains the following code:

```
#head(ext_tracks$hour) %>%
# select(storm_name, hour, max_wind) %>%
# filter(hour == "00") %>%
# head(3)

ext_tracks %>%
  group_by(storm_name, year) %>%
  summarize(worst_wind = max(max_wind)) %>%
  filter(worst_wind >= 160)
```

The console shows the execution of `select(storm_name, latitude, longitude, starts_with("radius_34"))` and the resulting tibble:

```
# A tibble: 2 x 3
  storm_name year worst_wind
  <chr>      <dbl>    <dbl>
1 GILBERT   1988      160
2 WILMA     2005      160
```

**Bottom Screenshot:** The script file `jaspas_lab3.R` contains the following code:

```
#select(storm_name, hour, max_wind) %>%
# filter(hour == "00") %>%
# head(3)

ext_tracks %>%
  group_by(storm_name, year) %>%
  summarize(worst_wind = max(max_wind)) %>%
  filter(worst_wind >= 160)

ext_tracks %>%
  select(storm_name, month, day, hour, latitude, longitude, max_wind) %>%
  filter(storm_name == "ANDREW" & max_wind >= 137)
```

The console shows the execution of `head(ext_tracks$hour)` and the resulting tibble:

```
# A tibble: 2 x 7
  storm_name month day hour latitude longitude max_wind
  <chr>      <chr> <chr> <chr>    <dbl>    <dbl>    <dbl>
1 ANDREW    08  23  12     25.4     74.2     145
2 ANDREW    08  23  18     25.4     75.8     150
```

## 5.) Adding, changing and Renaming columns

The first screenshot shows the initial data frame 'worldcup' being created and filtered. The second screenshot shows the addition of a new column 'ave\_shots' and the renaming of the 'Name' column to 'player\_name'. The third screenshot shows the final data frame with all columns and the resulting tibble output.

```
# First Screenshot: Initial Data Frame Creation and Filtering  
# ext_tracks %>%  
#   select(storm_name, month, day, hour, latitude, longitude, max_wind) %>%  
#   filter(storm_name == "ANDREW" & max_wind >= 137)  
worldcup <- worldcup %>%  
  mutate(player_name = rownames(worldcup))  
worldcup %>% slice(1:3)
```

The second screenshot shows the addition of a new column 'ave\_shots' and the renaming of the 'Name' column to 'player\_name'. The third screenshot shows the final data frame with all columns and the resulting tibble output.

```
# Second Screenshot: Adding and Renaming Columns  
worldcup <- worldcup %>%  
  group_by(Position) %>%  
  mutate(ave_shots = mean(Shots)) %>%  
  ungroup()  
worldcup %>% slice(1:3)  
worldcup <- worldcup %>%  
  rename(Name = player_name) %>%  
  slice(1:3)
```

The third screenshot shows the final data frame with all columns and the resulting tibble output.

```
# Third Screenshot: Final Data Frame and Output  
# A tibble: 3 x 9  
  Team      Position Time Shots Passes Tackles Saves Name ave_shots  
  <fct>    <fct>    <int> <int> <int> <int> <int> <chr> <dbl>  
1 Algeria Midfielder 16 0 6 0 0 0 Abdou 2.39  
2 Japan Midfielder 351 0 101 14 0 Abe 2.39  
3 France Defender 180 0 91 6 0 Abid 1.16
```



## 6.) Spreading and Gathering data

The image displays two screenshots of the RStudio interface, illustrating data manipulation tasks.

**Left Screenshot:** Shows an R script file named `jaspa_lab3.R`. The code includes comments and functions for selecting, filtering, mutating, and grouping data. The console output shows the result of a `head(VADeaths)` command, displaying a table of age groups and death counts.

```
#worldcup %>% slice(1:3)
#worldcup %>%
# rename(Name = player_name) %>%
# slice(1:3)
head(VADeaths)
```

```
> # select(storm_name, month, day, hour, latitude, longitude, max_wind) %>%
> # filter(storm_name == "ANDREW" & max_wind >= 137)
>
> #worldcup <- worldcup %>%
> # mutate(player_name = rownames(worldcup))
> #
> #worldcup %>% slice(1:3)
>
> #worldcup <- worldcup %>%
> # group_by(Position) %>%
> # mutate(ave_shots = mean(Shots)) %>%
> # ungroup()
> #
> #worldcup %>% slice(1:3)
>
> #worldcup %>%
> # rename(Name = player_name) %>%
> # slice(1:3)
>
> head(VADeaths)
```

	Rural	Male	Rural	Female	Urban	Male	Urban	Female
50-54	11.7	8.7	15.4	8.4				
55-59	18.1	11.7	24.3	13.6				
60-64	26.9	20.3	37.0	19.3				
65-69	41.0	30.9	54.6	35.1				
70-74	66.0	54.3	71.1	50.0				

**Right Screenshot:** Shows the same R script file, but with an error in the `mutate` function. The console output displays the error message: "Error in 'mutate()': object 'age' not found". The warning message indicates that `tbl_df()` is deprecated in dplyr 1.0.0 and should be replaced with `as_tibble()`.

```
#worldcup %>%
# rename(Name = player_name) %>%
# slice(1:3)
#head(VADeaths)
VADeaths <- VADeaths %>%
tbl_df() %>%
mutate(age = row.names(VADeaths))
VADeaths
```

```
> #worldcup %>%
> # rename(Name = player_name) %>%
> # slice(1:3)
>
> #head(VADeaths)
>
> VADeaths <- VADeaths %>%
+   tbl_df() %>%
+   mutate(age = row.names(VADeaths))
Error in `mutate()`:
! object 'age' not found
! In argument: `age = row.names(VADeaths)`.
Caused by error:
! object 'age' not found
! Run `rlang::last_trace()` to see where the error occurred.
Warning message:
! `tbl_df()` was deprecated in dplyr 1.0.0.
! Please use `tibble::as_tibble()` instead.
! Call `lifecycle::last_lifecycle_warnings()` to see where this
! warning was generated.
> VADeaths
```

	Rural	Male	Rural	Female	Urban	Male	Urban	Female
50-54	11.7	8.7	15.4	8.4				
55-59	18.1	11.7	24.3	13.6				
60-64	26.9	20.3	37.0	19.3				
65-69	41.0	30.9	54.6	35.1				
70-74	66.0	54.3	71.1	50.0				

# Adrian Joel Jaspa

## CMSC 124 - LAB

