Take Home Quiz 1

Your Name

Due Tuesday Feb. 6 at 12:45pm

This quiz should take you approximately 25 minutes. Place your answers into this markdown document, knit it, and hand in the result as a PDF or Word document. You may use R, any reference material, and information already available on the internet. Do not work together and do not get help, except from Dr. Clair.

Problem 1 (10 points)

Problems 1-3 use the `storms` data set which is in the `dplyr` library. Each row represents one observation of a hurricane on the Atlantic coast.

a. How variables are in this data?

b. How many rows are in this data?

c. According to the help page, what are the units on the `wind` and `pressure` variables?

Solution

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
str(storms)
```

```
## tibble [19,537 x 13] (S3: tbl_df/tbl/data.frame)
##  $ name                        : chr [1:19537] "Amy" "Amy" "Amy" "Amy" ...
##  $ year                        : num [1:19537] 1975 1975 1975 1975 1975 ...
##  $ month                       : num [1:19537] 6 6 6 6 6 6 6 6 6 6 ...
##  $ day                         : int [1:19537] 27 27 27 27 28 28 28 28 29 29 ...
##  $ hour                        : num [1:19537] 0 6 12 18 0 6 12 18 0 6 ...
##  $ lat                         : num [1:19537] 27.5 28.5 29.5 30.5 31.5 32.4 33.3 34 34.4 34 ...
##  $ long                        : num [1:19537] -79 -79 -79 -79 -78.8 -78.7 -78 -77 -75.8 -74.8 ...
##  $ status                      : Factor w/ 9 levels "disturbance",..: 7 7 7 7 7 7 7 7 8 8 ...
##  $ category                    : num [1:19537] NA NA NA NA NA NA NA NA NA NA ...
##  $ wind                        : int [1:19537] 25 25 25 25 25 25 25 30 35 40 ...
##  $ pressure                    : int [1:19537] 1013 1013 1013 1013 1012 1012 1011 1006 1004 1002 ...
##  $ tropicalstorm_force_diameter: int [1:19537] NA NA NA NA NA NA NA NA NA NA ...
##  $ hurricane_force_diameter    : int [1:19537] NA NA NA NA NA NA NA NA NA NA ...
```

13 variables, 19537 rows, wind in knots and pressure in millibars.

Problem 2 (10 points)

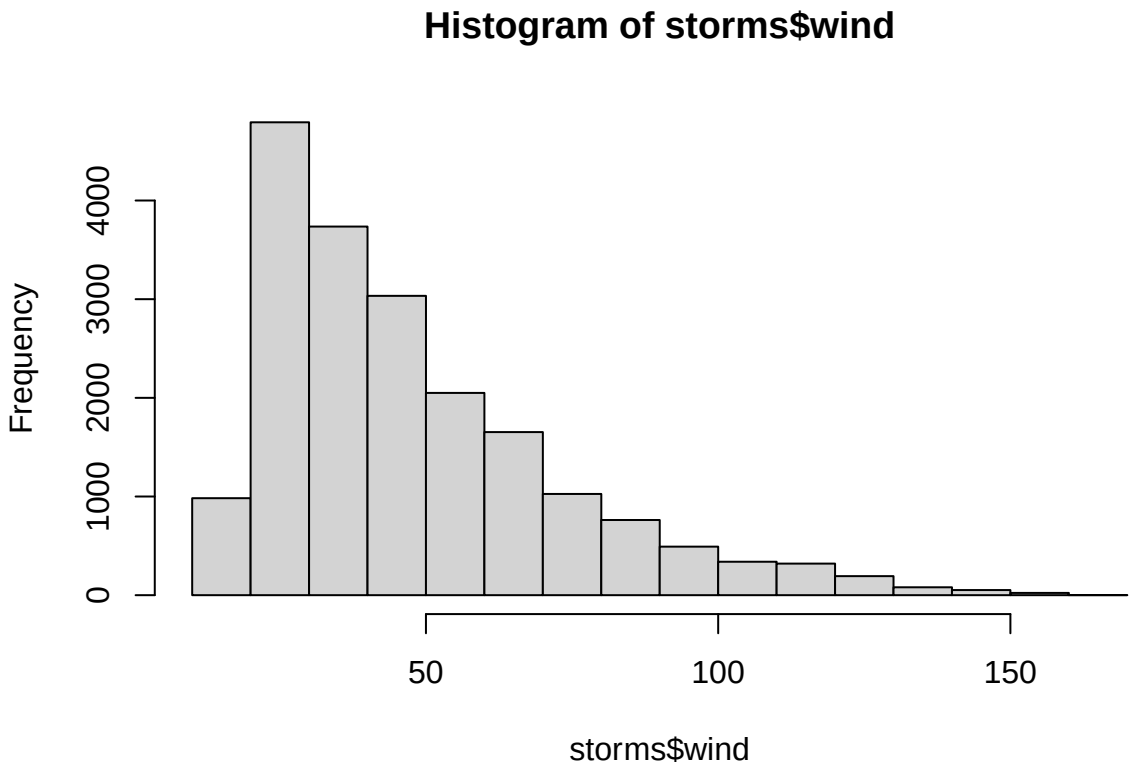
(continue using `storms`)

a. Make a histogram of wind speeds.

b. Which storm had the highest recorded wind speed? What year was that?

Solution

```
hist(storms$wind)
```




```
storms |> slice_max(wind)
```

```
## # A tibble: 1 x 13
##   name   year month   day  hour   lat  long status    category  wind pressure
##   <chr> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <fct>        <dbl> <int>    <int>
## 1 Allen  1980     8     7    18  21.8 -86.4 hurricane        5   165      899
## # i 2 more variables: tropicalstorm_force_diameter <int>,
## #   hurricane_force_diameter <int>
```

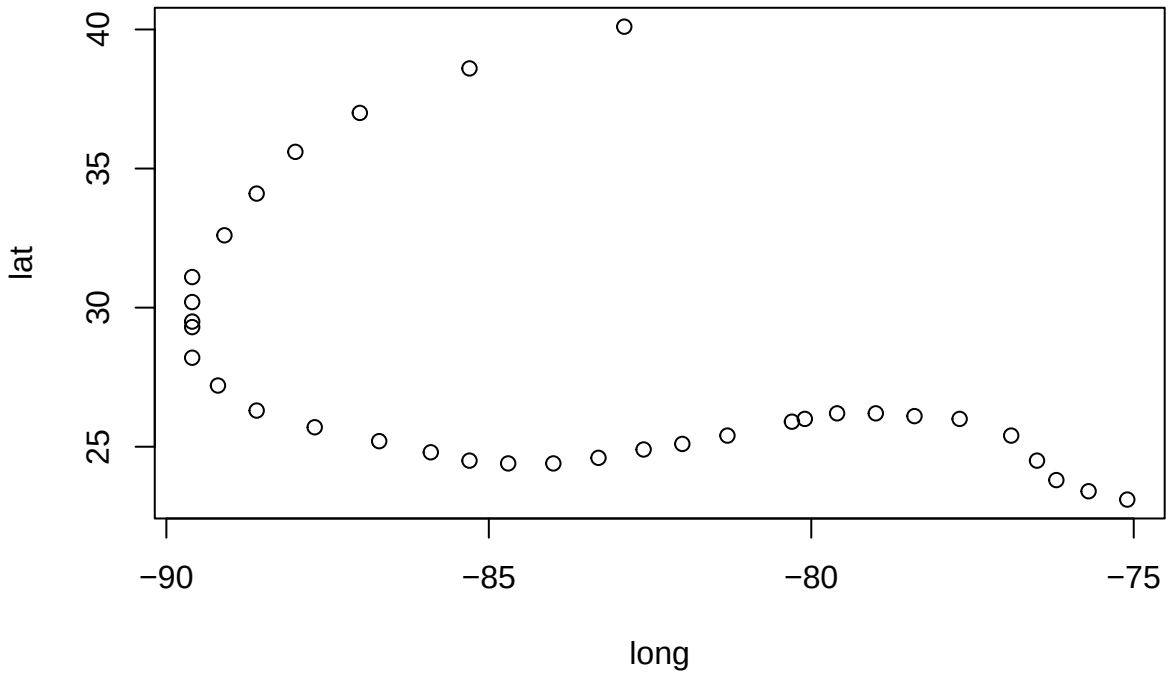
Problem 3 (10 points)

(continue using `storms`)

In 2005, Katrina hit New Orleans. Make a plot of the storm track for the 2005 Katrina by plotting `long` on the x-axis and `lat` on the y-axis.

You should see the storm hit New Orleans at approximately 30N and -90W.

```
storms |> filter(year==2005, name=="Katrina") |> select(long, lat) |> plot()
```



Problem 4 (10 points)

On episode 2 of the Nicolas Cage documentary, "The History of Swear Words," six comedians took part in an experiment. Each was asked to submerge their hand in a bucket of ice water for as long as they could stand it. Four of the comedians were allowed to swear out loud while performing the experiment, the other two were forbidden from swearing. The two clean subjects spent 53 and 58 seconds in the water, while the four swearing comedians spent 69, 78, 87, and 140 seconds in the water.

a. What is the difference in mean time between the swearing and non-swearing groups?

b. Perform a t-test for the difference in means. Report your conclusions with a p-value.

```
ice <- data.frame(swear = c('n','n','y','y','y','y'), icetime = c(53, 58, 69, 78, 87, 140))
t.test(icetime ~ swear, data=ice)
```

```
##
##  Welch Two Sample t-test
##
## data:  icetime by swear
## t = -2.3567, df = 3.1439, p-value = 0.09572
## alternative hypothesis: true difference in means between group n and group y is not equal to 0
## 95 percent confidence interval:
##  -88.01207  12.01207
## sample estimates:
## mean in group n mean in group y
##            55.5            93.5
```

The difference in means is 93.5-55.5 = 38. This difference is not significant (p = 0.096).

Problem 5 (10 points)

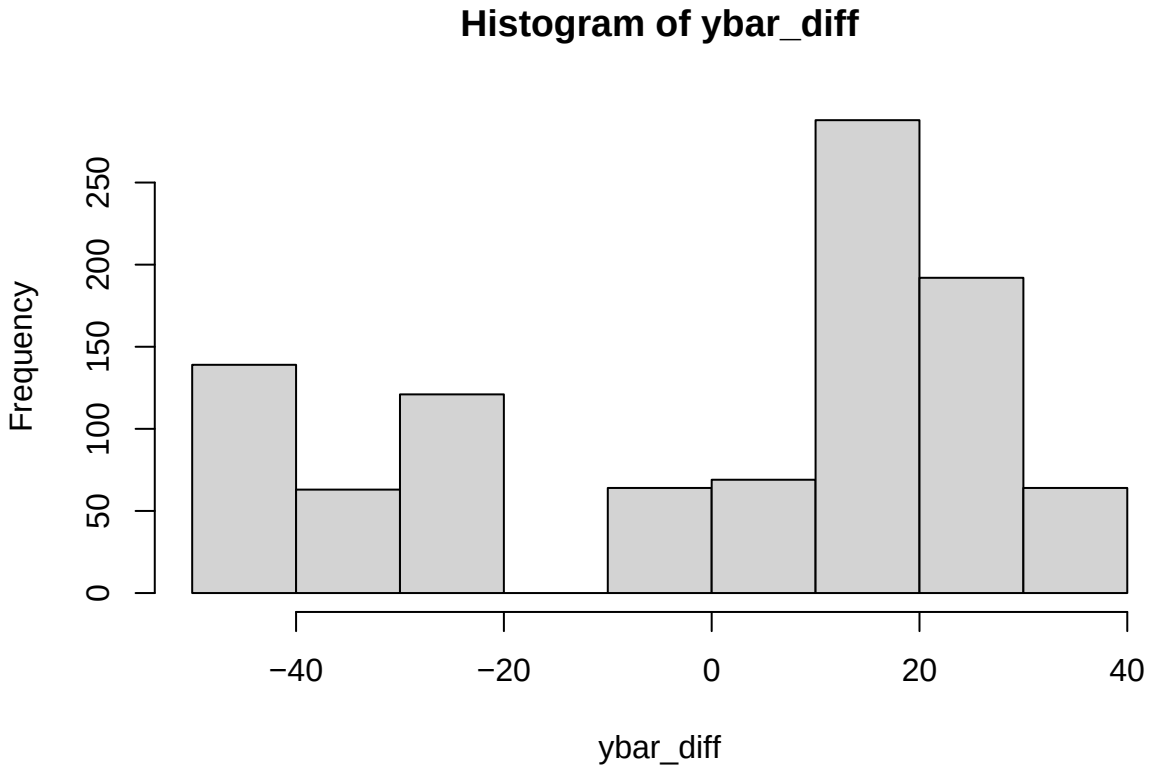
Continue with the History of Swear Words experiment.

a. Use resampling to make 1000 reassignments to groups. Compute the difference in means for each.

b. Make a histogram to show the sampling distribution of the difference in means.

c. Finish the randomization test by computing the probability that the difference in means is as large as the observed difference.

```
ybar_diff <- replicate(1000,
    ice |> mutate(swear = sample(swear)) |> group_by(swear) |>
      summarize(ybar = mean(icetime)) |> pull(ybar) |> diff()
)
hist(ybar_diff)
```




```
table(abs(ybar_diff) >= 38)
```

```
##
## FALSE  TRUE
##   797   203
```

```
mean(abs(ybar_diff) >= 38)
```

```
## [1] 0.203
```