Lab 05: Explore Other Distributions

```
library(tidyverse)
```

For this lab, you will conduct simulations related to the traffic light example started as group work in class.

Task 3

```
#use set.seed() to set a random seed

#conduct simulations you strategized in group work (Task 2)
```

Tasks 4 - 6

Do by hand. You can then (1) upload scanned document of your written work, (2) type up your written results using LaTeX, or (3) embed an image of your written work into this .qmd document.

Simulating Y

The following code simulates 10,000 values of the random variable described in Task 5, which we'll call `Y`, and plots the resulting distribution. We use $p = 0.75$ to mimic the traffic light example above.

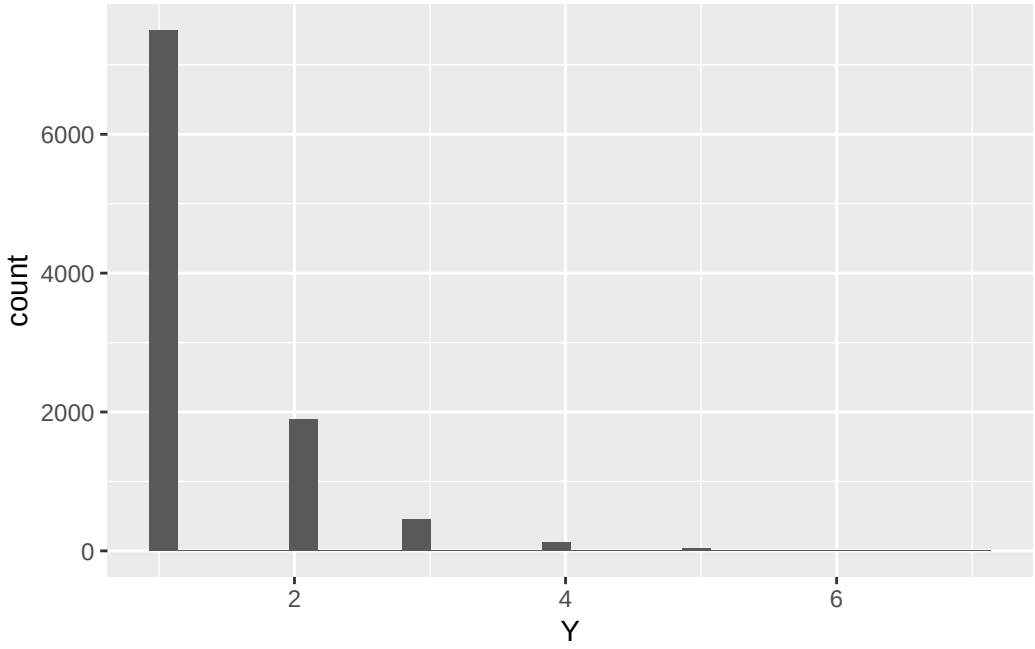
```
set.seed(437)
#initialize empty vector
#will fill with the # of trials needed to reach first success
Y <- c()
for(i in 1:10000){
```

```
  #initialize the # of successes and # of trials to be 0
  num_successes <- 0
  trials <- 0

  #while # of successes remains 0,
  #generate a new observation x of a bernoulli random variable
  #update the number of trials
  #once x = TRUE, num_successes will be updated to 1
  #and while loop will end
  while(num_successes < 1){
    x <- rbernoulli(1, .75)
    num_successes <- num_successes + x
    trials <- trials + 1
  }

  #fill element i of the vector Y with the number of trials
  #it took to reach the first success
  Y[i] <- trials
  #for loop repeats this process 10,000 times
}
```

```
Warning: `rbernoulli()` was deprecated in purrr 1.0.0.
```

```
#place simulation results in a dataframe
sims_Y <- data.frame(Y)

#plot simulation results
ggplot(sims_Y, aes(x = Y)) +
  geom_histogram()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Task 7

YOUR ANSWER HERE

Comparing Simulated and Theoretical Probabilities

We can calculate the theoretical probabilities using the pmf and compare them to the simulated probabilities.

```
#calculate the theoretical probabilities from the pmf
#create vector with all possible values of Y from the simulations
support_Y <- seq(1, max(Y))
support_Y
```

```
[1] 1 2 3 4 5 6 7
```

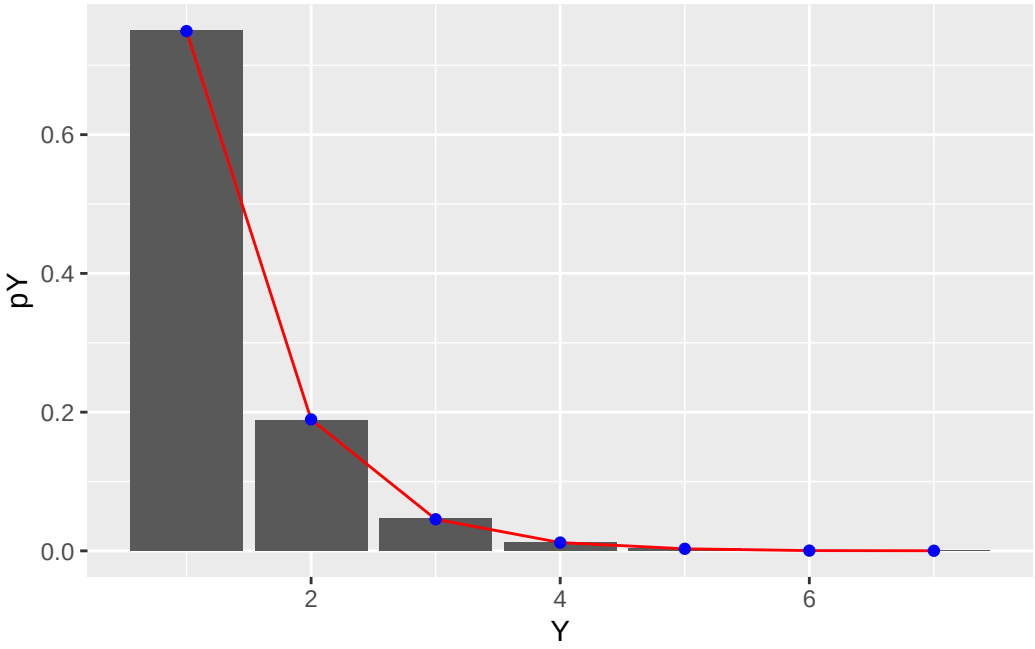
```
#probability of one success and y-1 failures
#calculated for each value of y in the support
pY <-  .75*(.25)^(support_Y-1)
pY
```

```
[1] 0.7500000000 0.1875000000 0.0468750000 0.0117187500 0.0029296875
[6] 0.0007324219 0.0001831055
```

```r
prob_dist_Y <- data.frame(Y = support_Y,
                          pY)

#add a column with the simulated relative frequencies
prob_dist_Y$pY_sims <- proportions(table(sims_Y))

prob_dist_Y
```

```
  Y           pY pY_sims
1 1 0.7500000000  0.7493
2 2 0.1875000000  0.1895
3 3 0.0468750000  0.0456
4 4 0.0117187500  0.0120
5 5 0.0029296875  0.0030
6 6 0.0007324219  0.0004
7 7 0.0001831055  0.0002
```

```r
#plot the theoretical probabilites on top of the simulated distribution
ggplot(prob_dist_Y) +
  geom_col(aes(x = Y, y = pY)) +
  geom_line(aes(x = Y, y = pY_sims), color = "red") +
  geom_point(aes(x = Y, y = pY_sims), color = "blue")
```



Task 8

REFLECTION PROMPTS

Respond to each of the following after you have completed all tasks in this activity

(RP1): What were the main concepts covered in this activity?

YOUR REFLECTION HERE

(RP2): What's one thing you understand better after completing these tasks?

YOUR REFLECTION HERE

(RP3): What task(s) gave you the most trouble? What was difficult about them/where did you get stuck?

YOUR REFLECTION HERE

SUBMISSION

Tasks 4 - 6 (done by hand) can either be typed up in this document using LaTeX, or you can scan your work and (1) submit it as a separate document or (2) embed it as an image in the Tasks 4 - 6 section above (when in Visual view, click Insert > Figure/Image).