# Example Paper

*Ryan Safner*

*11/29/2018*

**Abstract**

In this paper, I use fake data I created to demonstrate how to organize your files and manage your workflow effectively.

## Introduction

I am managing all of my files (paper, references, data, code, and figures) by creating an `R Project`. This sets my working directory `wd()` to a folder on my computer where the `R Project` is stored. Everything that I put within this folder (including folders, for my Data, my Figures, etc.) is accessible from the same starting directory. I can send the entire project to you, and you can use all of the content easily, since all files are referenced *relative* to the project's directory folder.

## Literature Review

I use the citekeys (beginning with `@`) from my `bibexample.bib` file in the **Bibliography** folder which contains all of my references, any time I wish to cite a reference in my text.

> "Here is an example quote from this article." Smith and Jones (2018, 2)

Doe, Gibberish, and Fakerson (2016, 12) says "this short quote", disagreeing with Smith and Jones (2018).

See also the references mentioned appear at the end of the document now (I just need to add a `# References` or `# Bibliography` at the end of the file to create a section for them). It lists all the references you cited in alphabetical order.

## Data Creation

I first created my random data with the script `01-generate-data.R` in my **Scripts** folder. This produces two files, `rawdata1.csv` and `rawdata2.csv`,[1] which I saved in my **Data** folder. My second script, `02-data-cleaning.R` takes the two data files and merges them in to a single dataset, which I save as `clean_data.csv` in my **Data** folder for later keeping. I can then send out `clean_data.csv` to coauthors and spectators who wish to use my data, but I can always justify where it came from because running `02-data-cleaning.R` on the two raw data sources, `rawdata1.csv` and `rawdata2.csv` will always produce `clean_data.csv` for anyone who runs it!

[1] This is to show that often we will need to import multiple data files and wrangle them into a tidy format for data analysis!

## Summary Statistics

Here are some summary statistics of my data, the script for which is saved in **Scripts** as `03-summary-statistics.R`. I reproduce the script as `R` chunks in my `R Markdown .Rmd` document to demonstrate that everything *could* be produced in one document. I hide the raw code by setting `echo=FALSE` in the chunk options, since I do not wish to display code in my final paper.

| Variable | Obs | Min | Q1 | Median | Q3 | Max | Mean | Std. Dev. |
|---|---|---|---|---|---|---|---|---|
| x | 500 | 1.99 | 4.27 | 4.93 | 5.69 | 8.81 | 4.97 | 1.06 |
| y | 500 | -31.13 | -6.61 | 0.85 | 8.25 | 31.56 | 0.76 | 10.44 |

| Var1 | Freq |
|---|---|
| circle | 149 |
| rectangle | 109 |
| square | 121 |
| triangle | 121 |

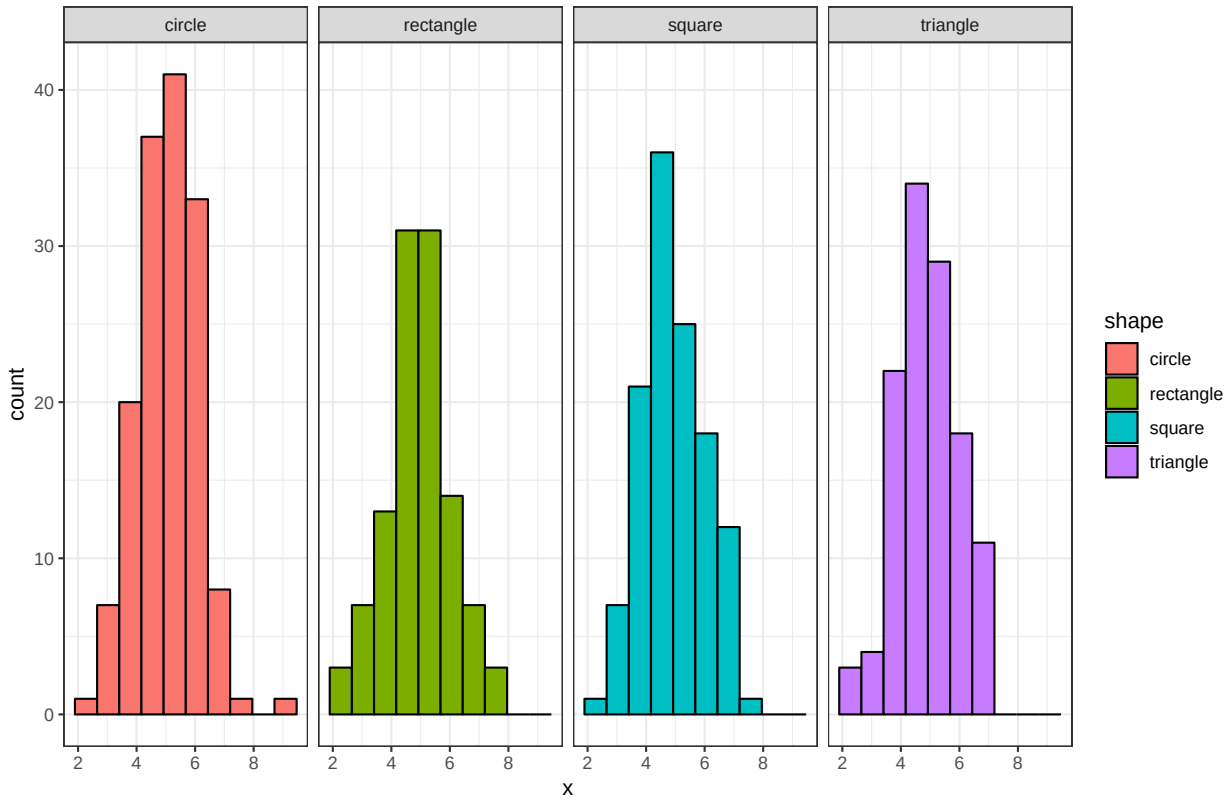


Figure 1: Histogram of x

# Plots

Here, I load as images from my **Figures** folder some of the plots that I generated in my `04-plots.R` script in my **Scripts** folder.

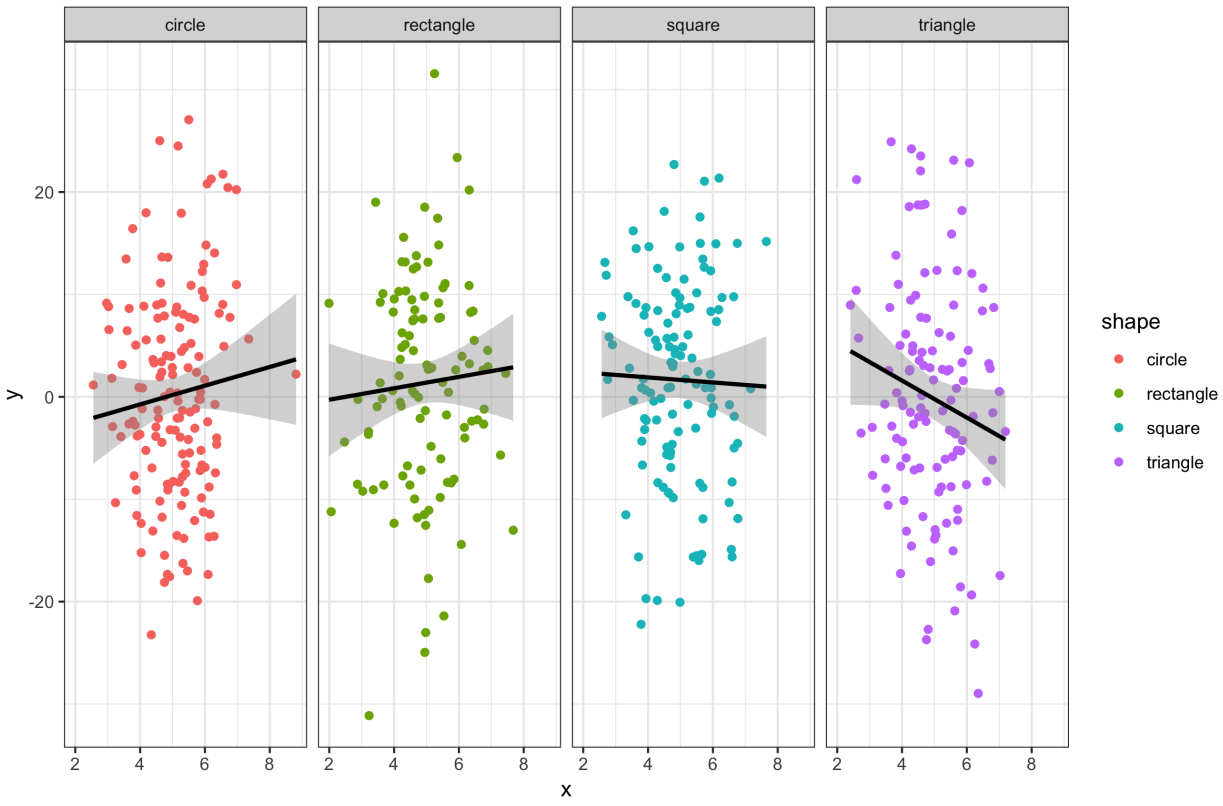


Figure 2: Scatterplot of x and y

# Regressions

I run three regressions with my fake data. The first is a simple regression of $y$ on $x$:

$$Y_i = \beta_0 + \beta_1 X_i$$

The second is augmented by including a dummy variable for each of the four categories of `shape` (omitting triangle).

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 circle_i + \beta_3 rectangle_i + \beta_4 square_i$$

The third adds "state-fixed effects" using the Least Squares Dummy Variable method.[2]

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 circle_i + \beta_3 rectangle_i + \beta_4 square_i + \alpha_i$$

All of the regressions are run in the script `05-regressions.R` in the **Scripts** folder. I import the script and use an `R` chunk below to run `stargazer` to generate the regression table.

| | *Dependent variable:* | | |
|---|---|---|---|
| | y | | |
| | | | With State-effects |
| | (1) | (2) | (3) |
| x | −0.130 | −0.110 | 0.019 |
| | (0.443) | (0.444) | (0.472) |
| shaperectangle | | 1.076 | 0.833 |
| | | (1.320) | (1.400) |
| shapesquare | | 1.401 | 0.725 |
| | | (1.282) | (1.379) |
| shapetriangle | | −0.332 | −0.609 |
| | | (1.282) | (1.368) |
| Constant | 1.402 | 0.809 | −2.312 |
| | (2.248) | (2.415) | (4.521) |
| Observations | 500 | 500 | 500 |
| $R^2$ | 0.0002 | 0.005 | 0.086 |
| Adjusted $R^2$ | −0.002 | −0.003 | −0.025 |
| Residual Std. Error | 10.449 (df = 498) | 10.456 (df = 495) | 10.568 (df = 445) |
| F Statistic | 0.086 (df = 1; 498) | 0.605 (df = 4; 495) | 0.776 (df = 54; 445) |
| *Note:* | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

[2]To avoid showing all of the state dummies, I add `omit=c("state")` to the `stargazer()` command.

## References

Doe, John, Erica Gibberish, and Frank Fakerson. 2016. *How Did This Get Published?* New York, NY: Sketchy Publishing, Inc.

Smith, Robert, and Anne Jones. 2018. "A Test Article." *Fake Articles Quarterly* 12 (2): 1–21.