Tips-Example

Susan VanderPlas

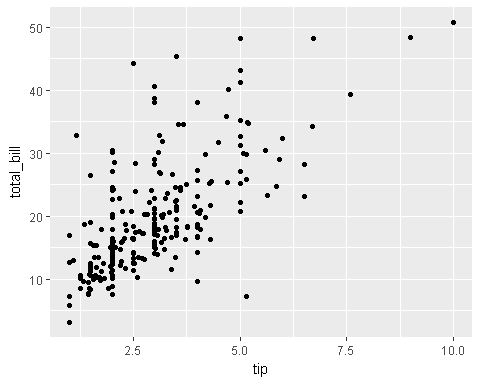
March 21, 2017

## Data Exploration

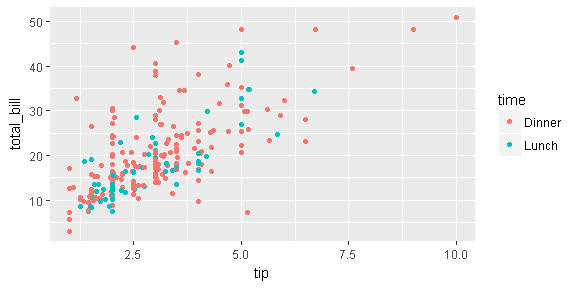
This dataset has a few rows and columns. Here's what it looks like:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| total\_bill | tip | sex | smoker | day | time | size |
| 16.99 | 1.01 | Female | No | Sun | Dinner | 2 |
| 10.34 | 1.66 | Male | No | Sun | Dinner | 3 |
| 21.01 | 3.5 | Male | No | Sun | Dinner | 3 |
| 23.68 | 3.31 | Male | No | Sun | Dinner | 2 |
| 24.59 | 3.61 | Female | No | Sun | Dinner | 4 |
| 25.29 | 4.71 | Male | No | Sun | Dinner | 4 |

## Graphical Exploration

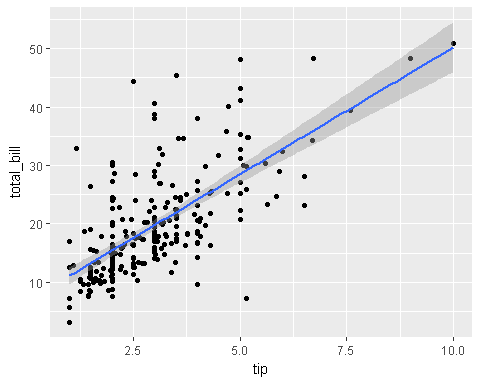
Here is a basic scatterplot of tips vs. total bill 

Lalalala useful information



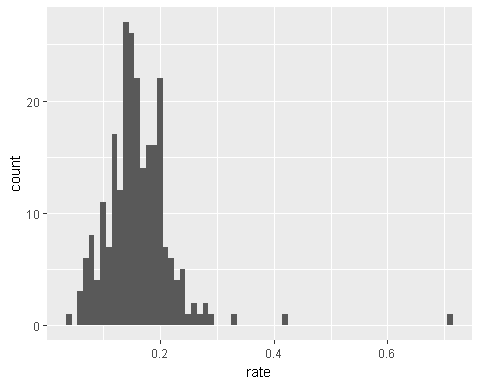
Scatterplot in full color

Linear regressions are awesome. Try them!



## Creating new variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 0.03564 | 0.1291 | 0.1548 | 0.1608 | 0.1915 | 0.7103 |



Someone is a really awesome tipper! Who are they?

## total\_bill tip sex smoker day time size rate  
## 173 7.25 5.15 Male Yes Sun Dinner 2 0.7103448

Welch Two Sample t-test: rate by sex

|  |  |  |  |
| --- | --- | --- | --- |
| Test statistic | df | P value | Alternative hypothesis |
| 1.143 | 206.8 | 0.2542 | two.sided |