David Schlosz Assignment 2

­­­­­­­­­

1. After searching numerous databases, I eventually found the following:

<https://raw.githubusercontent.com/fivethirtyeight/data/master/tarantino/tarantino.csv>

This is a list of Tarantino movies and the number of swear words in each movie. This seemed entertaining to work with.

I imported this file directly into R studio and called it T2

1. Variables:

“movie”, “type”, “word”, “minutes in”

1. I created t2a by removing “type” and “minutes in”. This left two remaining variables: “movie” and “word”.
2. Because there were too many swear words, I created a value called “not words”. Which, when printed in table form, looks as follows:

word

movie asshole dick fuck hell shit

Django Unchained 47 0 0 8 10 17

Inglorious Basterds 48 0 0 11 4 3

Jackie Brown 4 0 1 44 10 66

Kill Bill: Vol. 1 63 0 0 5 3 8

Kill Bill: Vol. 2 11 1 0 3 1 6

Pulp Fiction 7 3 0 74 5 76

Reservoir Dogs 10 6 13 68 12 45

1. Data Structure: Variables in this the t2a dataset are: “movie” and “word”

I wanted to make each of the movies a variable but couldn’t remember or figure that out.

1. Mean:
   1. Asshole: 2.25
   2. Dick 3.25
   3. Fuck 40.25
   4. Hell 7.75
   5. Shit 35.25
2. Mode: Could not figure this out.
3. Median:
   1. Asshole: 1.5
   2. Dick 0
   3. Fuck 39.5
   4. Hell 7.5
   5. Shit 31
4. SD:
   1. Asshole: 2.872281
   2. Dick 6.5
   3. Fuck 35.6125
   4. Hell 3.86221
   5. Shit 32.29422
5. Variance:
   1. Asshole: na/Nan argument ???
   2. Dick 42.25
   3. Fuck 1268.25
   4. Hell 14.91667
   5. Shit 1042.917
6. Range:
   1. Asshole: min/max 0 6 range: 6
   2. Dick min/max 0 13 range: 13
   3. Fuck min/max 8 74 range: 66
   4. Hell min/max 4 12 range: 8
   5. Shit min/max 3 76 range: 73
7. Stem:

Asshole: 0 | 00

2 | 0

4 |

6 | 0

1. Histogram:

