

Analysis using Python and R

SCRUB A DATA SET OF 5000+ LINES OF TEXT AND PERFORM STATISTICAL ANALYSIS on the data

Roisin Carroll | Programming for Big Data B8IT105 | 7-08-17

Contents

[Statistical Analysis Conclusions 2](#_Toc489893895)

[1. Active Days 2](#_Toc489893896)

[2. Commits by Author 3](#_Toc489893897)

[3. Commits by Type 4](#_Toc489893898)

# Statistical Analysis Conclusions

## Active Days

From analyzing at the number of commits by day, I can see that;

1. There are no commits on either Saturdays or Sundays as these are days off.
2. Thursday appears to be the busiest day for the development team, while Monday appears to be the quietest day.
3. Commits generally increase as the week goes on except for Wednesday. Perhaps there is a meeting occupying developers time that day.

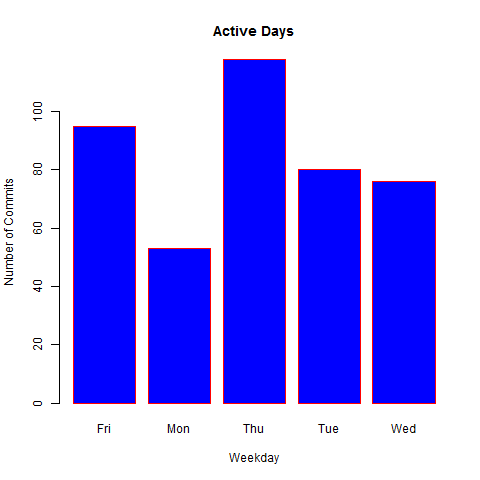
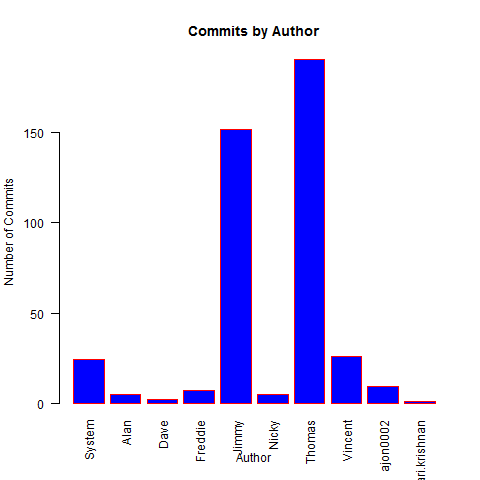


Fig 1: Bar chart of Active Days

## Commits by Author

From analyzing at the total number of commits by author, I can see that;

1. Thomas and Jimmy appear to be the most productive developers within the team since they have committed the most changes to the codebase.
2. There is a massive disparity between the developers in terms of commits. This could be for a number of reasons. Perhaps the other team-members are being used more in meetings/admin/managerial work and less in coding, or they could be new/junior members. Further investigation would be required before drawing conclusions, especially if productivity needs to be higher than it currently is.



## Commits by Type

From analyzing the type of commits I can see that;

1. There are more modifications than either additions or deletions by a large margin. This suggests that a lot of work is being done on a current piece of work. Perhaps there are some issues or problems with the code that need solving.

