Informedb Data Markdown File

Liam Whitenack

2021-05-27

This project was done as an Internship Report to be presented at the end of the year. It is made in an attempt to find the eight Software Quality Metrics found in the PowerPoint attatched.

**All data is collected from the Informedb project. The project has assigned issue resolution data on Jira and Has commit data on GitLab that was collected and organized to analyze in RStudio.**

## Question of Interest

The issue I will be addressing will be the military’s request for organized and presented data on projects. Software Quality Metrics are useful for analyzing trends and discovering patterns that lead to productivity. This project is being developed because the military would like reports on these metrics but doing some data analysis could show how the company can improve the efficiency of their software development.  
I will be addressing the issue by mining all the data available on Jira and GitHub for the Informedb Project on the company’s progress in finding issues, starting work on said issues, and resolving the issues. The project will require me to find eight different Software Quality Metrics:  
• Direct Trends  
• Total Defects found in the last 4 weeks  
• MTTR – Mean time to repair  
• Direct Removal Efficiency  
• Number of failed fix attempts (external)  
• Number of failed fix attempts (internal)  
• Change failure rate within release cycle  
• Automated Code Coverage (in percent)

## Preprocessing

GitLab\_bigger <- GitLab %>%  
 mutate(  
 a\_unix = `Author Date (Unix Timestamp)`,  
 a\_date = `Author Date`,  
 a\_name = `Author Name`,  
 a\_email = `Author Email`,  
 c\_unix = `Comitter Date (Unix Timestamp)`,  
 c\_date = `Committer Date`,  
 c\_name = `Committer Name`,  
 c\_email = `Committer Email`  
 )

Jira\_bigger <- Jira %>%  
 mutate(  
 priority = recode(  
 Priority,  
 `1 - Trivial` = "R1",  
 `2 - Minor` = "R1",  
 `3 - Major` = "R2",  
 `4 - Critical` = "R3",  
 `5 - Blocker` = "R4"  
 ),  
 time\_spent = `Time Spent (hours)`,  
 summary = Summary,  
 parent = `Parent id`,  
 issue = `Issue id`,  
 type = `Issue Type`,  
 date\_created = Created,  
 unix\_created = `Created (in Unix)`,  
 date\_updated = `Updated`,  
 unix\_updated = `Updated (in Unix)`,  
 date\_resolved = `Resolved`  
 )

GitLab\_reduced <- GitLab\_bigger %>%  
 select(`a\_date`, `a\_name`, `a\_unix`, `c\_date`, `c\_name`, `c\_unix`, Subject)

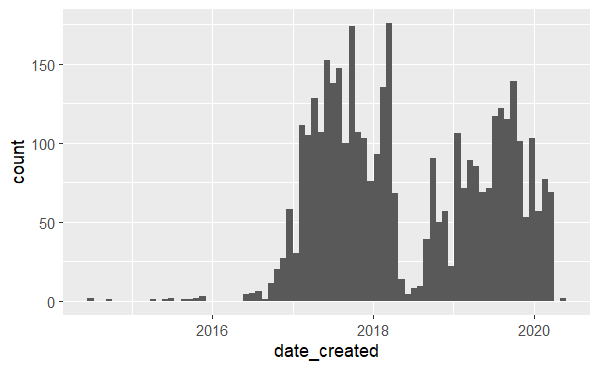
Jira\_reduced <- Jira\_bigger %>%  
 select(priority, time\_spent, date\_created, date\_updated, date\_resolved, unix\_created, unix\_updated, parent, type, Assignee, Reporter, Status, Summary)

This code block renames and selects the columns that will be used for practicality.

## Visualization

The project board has not been used since last March, but the distribution of

Jira\_reduced %>%  
 ggplot() +   
 geom\_histogram(aes(x = date\_created), bins = 77)



## Summary Statistics

Jira\_reduced %>%  
 group\_by(priority) %>%  
 summarise(   
 Count = n(),  
 )

## # A tibble: 4 x 2  
## priority Count  
## <chr> <int>  
## 1 R1 3600  
## 2 R2 94  
## 3 R3 68  
## 4 R4 73

\*\* Total Defects found in the last 4 weeks: The Informedb Jira page has not been used over the last 4 weeks, so I cannot tell how many issues were found. When the data was being recorded, there were a total of 3862 total issues found. 3600 were R1, 94 were R2, 68 were R3, and 73 were R4. That means that the mean number of each issues found over a four week span would be R1-46.86, R2-1.22, R3-0.89, R4-0.95. **MTTR: The average amount of time spent on each issue is 2250.65 hours (94 days), but many of the R1 issues are abandoned or not updated. The average spent from creation of an issue to its resolution is 1333.86 hours (56 days). The mean time to repair each issue in days would be R1-94, R2-31, R3-62, R4-143.**

## Data Analysis

## Conclusion