# Informedb Data Markdown File

Liam Whitenack

2021-06-01

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 <- GitLab %>%
  mutate(
    author_date = as.Date(strptime(author_date, format = "%a %b %d %H:%M:%S %Y")),
    committer_date = as.Date(strptime(committer_date, format = "%a %b %d %H:%M:%S %Y"))
  )
```

```
GitLab_reduced <- GitLab %>%
  select(
    author_date,
    committer_date
  )
```

```
Jira <- Jira %>%
    mutate(
      priority = recode(
       Priority,
          `1 - Trivial` = "R1",
          `2 - Minor` = "R1",
          `3 - Major` = "R2",
          `4 - Critical` = "R3",
          `5 - Blocker` = "R4"
      ),
      completed = recode(
        Status,
        Resolved = 1,
        Done = 1,
        Closed = 0,
        `In Progress` = 0,
        `Ready For Review` = 0,
        `Ready For Test` = 0,
        Reopened = 0,
        .default = 0
      ),
      Parent = `Parent id`,
      Issue = `Issue id`,
      Type = `Issue Type`,
      date_created = as.Date(strptime(Created, format = "%m/%d/%Y %H:%M")),
      date_updated = as.Date(strptime(Updated, format = "%m/%d/%Y %H:%M")),
      date_resolved = as.Date(strptime(Resolved, format = "%m/%d/%Y %H:%M")),
      time_spent = date_updated - date_created,
      repair_time = date_resolved - date_created
    )
Jira_reduced <- Jira %>%
  select(
    priority,
    completed,
    date_created,
    date_updated,
    date_resolved,
    time_spent,
    repair_time
  )
```

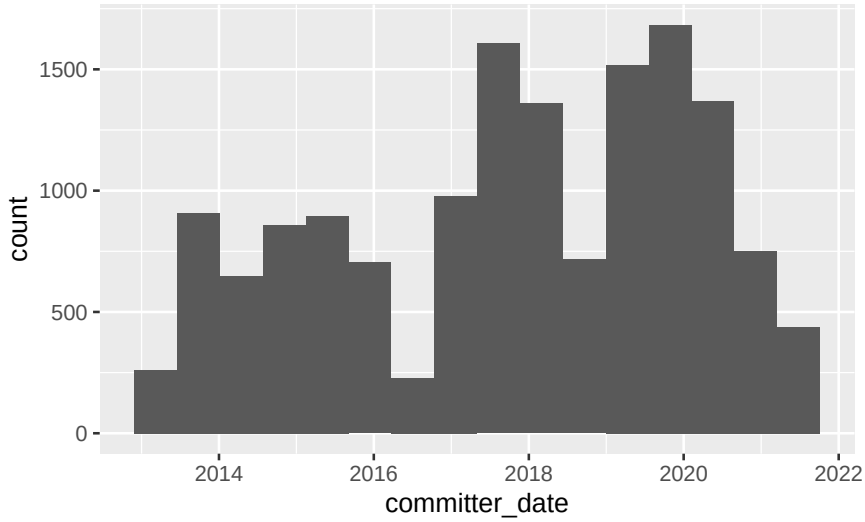
This code block renames and selects the columns that will be used for practicality.

### Visualization

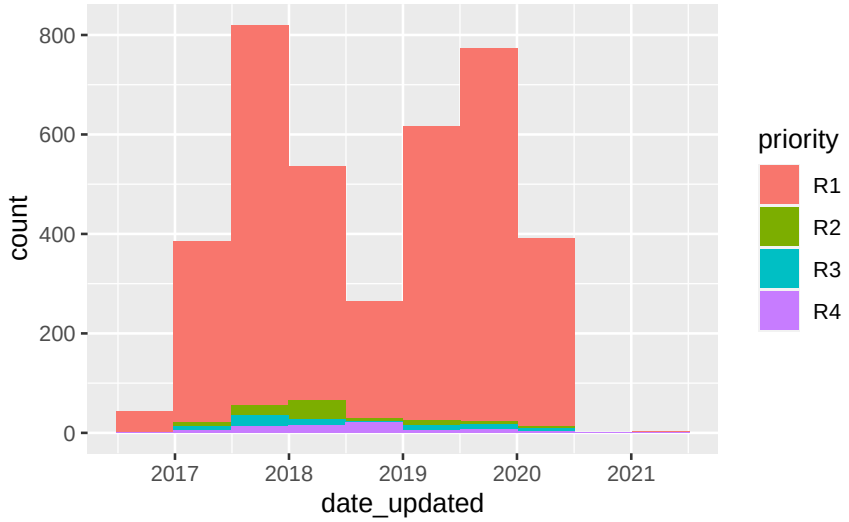
### Summary Statistics

```
Number_of_weeks <- (max(Jira_reduced$date_updated) - min(Jira_reduced$date_created))/7
Number_of_weeks <- as.numeric(Number_of_weeks)
```

```
GitLab_reduced %>%
  ggplot() +
  geom_histogram(aes(x = committer_date), bins = 16)
```




```
Jira_reduced %>%
  ggplot() +
  geom_histogram(aes(x = date_updated, fill = priority), bins = 10)
```




```
Jira_reduced %>%
  group_by(priority) %>%
  summarise(
    defects = n(),
    defects_in_four_weeks = n() / (Number_of_weeks / 4),
    MTTR_days = (mean((repair_time), na.rm = TRUE)),
    DRE = sum(completed) / n()
  )
```

| priority | defects | defects_in_four_weeks | MTTR_days | DRE |
|---|---|---|---|---|
| R1 | 3600 | 39.8891967 | 66.94241 days | 0.4391667 |
| R2 | 94 | 1.0415512 | 104.55556 days | 0.4574468 |
| R3 | 68 | 0.7534626 | 72.00000 days | 0.5441176 |
| R4 | 73 | 0.8088643 | 122.03333 days | 0.6575342 |

## Data Analysis

## Conclusion