

Case Study Report

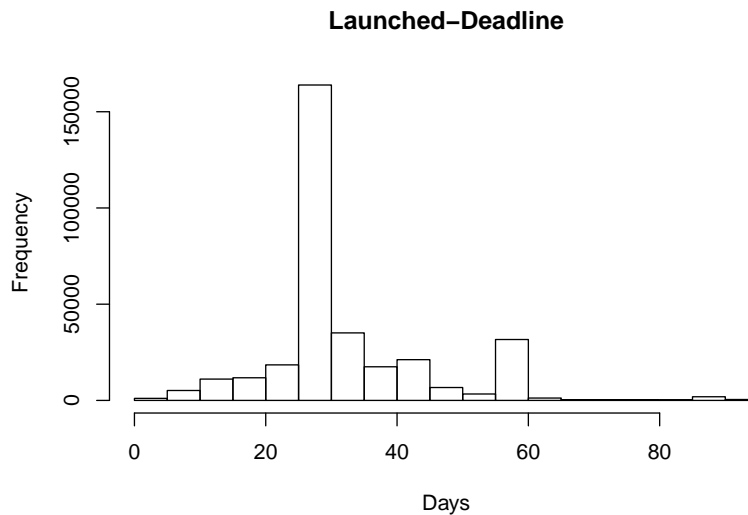
Colin Pi, Sharan Ganjam Seshachallam

2019 2 15

```
kick.2018 <- read_csv("ks-projects-201801.csv")
```

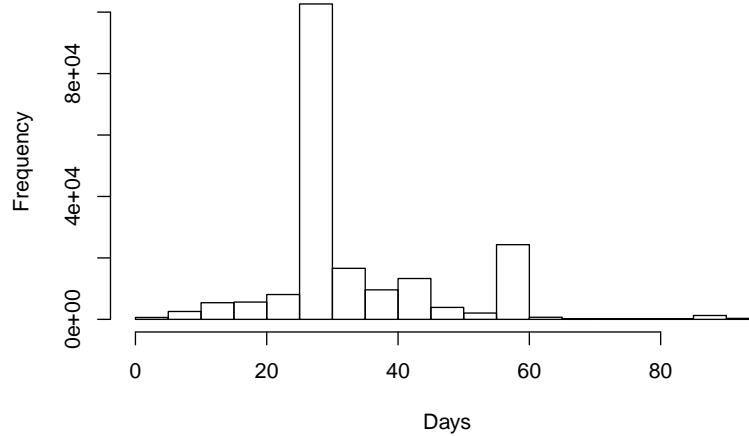
Launched-Deadline and Success/Failure

```
kick.2018 <- kick.2018 %>% filter(state %in% c("failed", "successful")) %>%  
  mutate(diff_date = as.numeric(as.Date(deadline) - as.Date(str_extract(launched,  
    "~.{10}"))))  
hist(kick.2018$diff_date, main = "Launched-Deadline", xlab = "Days")
```



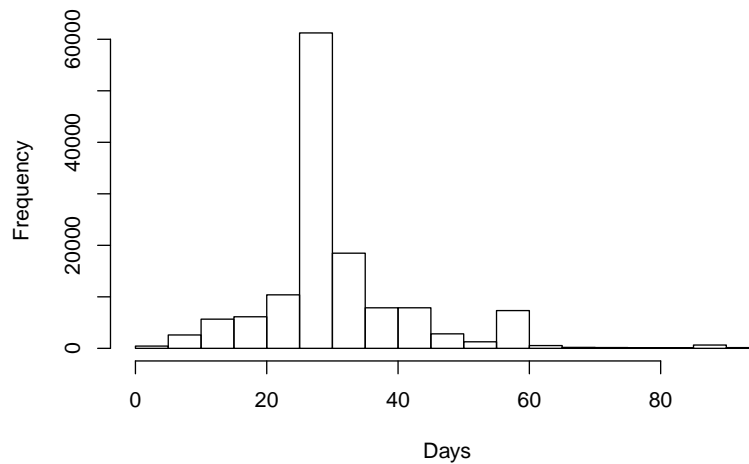
```
kick.2018$state <- factor(kick.2018$state)  
  
failed <- subset(kick.2018, select = diff_date, subset = state == "failed",  
  drop = T)  
successful <- subset(kick.2018, select = diff_date, subset = state == "successful",  
  drop = T)  
  
hist(failed, main = "Launched-Deadline, Failed", xlab = "Days")
```

Launched-Deadline, Failed



```
hist(successful, main = "Launched-Deadline, Successful", xlab = "Days")
```

Launched-Deadline, Successful



```
t.test(kick.2018$diff_date ~ kick.2018$state)
```

```
##
##  Welch Two Sample t-test
##
## data:  kick.2018$diff_date by kick.2018$state
## t = 68.989, df = 307680, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.931166 3.102584
## sample estimates:
##      mean in group failed mean in group successful
##           35.17332           32.15645
```

We are 95% confident that in average the days from launched date to deadline of the failed projects is from 2.93 to 3.10 days longer than that of the successful projects.

Success Rate before and after 2017

```
kick.2018 <- kick.2018 %>% mutate(before_2017 = ifelse(launched < "2017-01-01",
  "Yes", "No"))

before.2017 <- subset(kick.2018, select = state, subset = before_2017 == "Yes",
  drop = T)
after.2017 <- subset(kick.2018, select = state, subset = before_2017 == "No",
  drop = T)

succ.before.2017 <- sum(before.2017 == "successful")
succ.after.2017 <- sum(after.2017 == "successful")

prop.test(c(succ.before.2017, succ.after.2017), c(length(before.2017), length(after.2017)),
  correct = FALSE)

##
## 2-sample test for equality of proportions without continuity
## correction
##
## data: c(succ.before.2017, succ.after.2017) out of c(length(before.2017), length(after.2017))
## X-squared = 93.689, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.02943464 -0.01946908
## sample estimates:
##   prop 1   prop 2
## 0.4006772 0.4251290
```

We are 95% confident that in average the proportion of successful projects after 2017 is 2 to 3% higher than before 2017.