RQ2: Does the interest differ based on the type of TD?

Yasutaka Kamei

Feb 11th, 2016

Read Data

```
setwd("/Users/kamei/Research/techdebt/msr16_td_interest/")
source("./r_scripts/data_read.r")
```

The number of technical debt in each type

```
b <- data.CountLine.all

tmp <- b[b[,"Project"] == "apache-ant", ]
fc <- factor(tmp$Type)
tapply(tmp$interest, fc, length)
```

```
##      DEFECT      DESIGN REQUIREMENT        TEST
##           9          52           5           5
```

```
tapply(tmp$interest, fc, function(x){length(x)/length(tmp$interest) * 100})
```

```
##      DEFECT      DESIGN REQUIREMENT        TEST
##   12.676056   73.239437    7.042254    7.042254
```

```
tmp <- b[b[,"Project"] == "apache-jmeter", ]
fc <- factor(tmp$Type)
tapply(tmp$interest, fc, length)
```

```
##        DEFECT        DESIGN DOCUMENTATION   REQUIREMENT          TEST
##            11           148             2            16             4
```

```
tapply(tmp$interest, fc, function(x){length(x)/length(tmp$interest) * 100})
```

```
##        DEFECT        DESIGN DOCUMENTATION   REQUIREMENT          TEST
##      6.077348     81.767956      1.104972      8.839779      2.209945
```

```
tmp <- b[b[,"Project"] == "jruby", ]
fc <- factor(tmp$Type)
tapply(tmp$interest, fc, length)
```

```
##        DEFECT        DESIGN DOCUMENTATION   REQUIREMENT          TEST
##            87           108             1            37             3
```

```
tapply(tmp$interest, fc, function(x){length(x)/length(tmp$interest) * 100})
```

```
##        DEFECT        DESIGN DOCUMENTATION   REQUIREMENT          TEST
##    36.8644068    45.7627119     0.4237288    15.6779661     1.2711864
```

Observation

- In jruby, three types of technical dept are included than 10%.
- We use DEFECT, DESIGN and REQ in jruby.

Does the interest differ based on the type of TD?

```
tmp <- b[b[,"Project"] == "jruby", ]
tmp <- tmp[(tmp[,"Type"] == "DEFECT" |  tmp[,"Type"] == "DESIGN" |  tmp[,"Type"] == "REQUIREMENT"),]
fc <- factor(tmp$Type)

tapply(tmp$interest, fc, function(x){c(sum(x==0)/length(x)*100)})
```

```
##      DEFECT      DESIGN REQUIREMENT
##    52.87356    33.33333    45.94595
```

```
tapply(tmp$interest, fc, summary)
```

```
## $DEFECT
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -83.780   0.000   0.000  10.370   8.402 200.000
##
## $DESIGN
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -95.830  -4.944   0.000   4.466   7.853 362.500
##
## $REQUIREMENT
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -64.710  -2.941   0.000   2.912   0.000  64.440
```

```
defect <- subset(tmp$interest, tmp[,"Type"] == "DEFECT")
design <- subset(tmp$interest, tmp[,"Type"] == "DESIGN")
req <- subset(tmp$interest, tmp[,"Type"] == "REQUIREMENT")

library(vioplot)
```

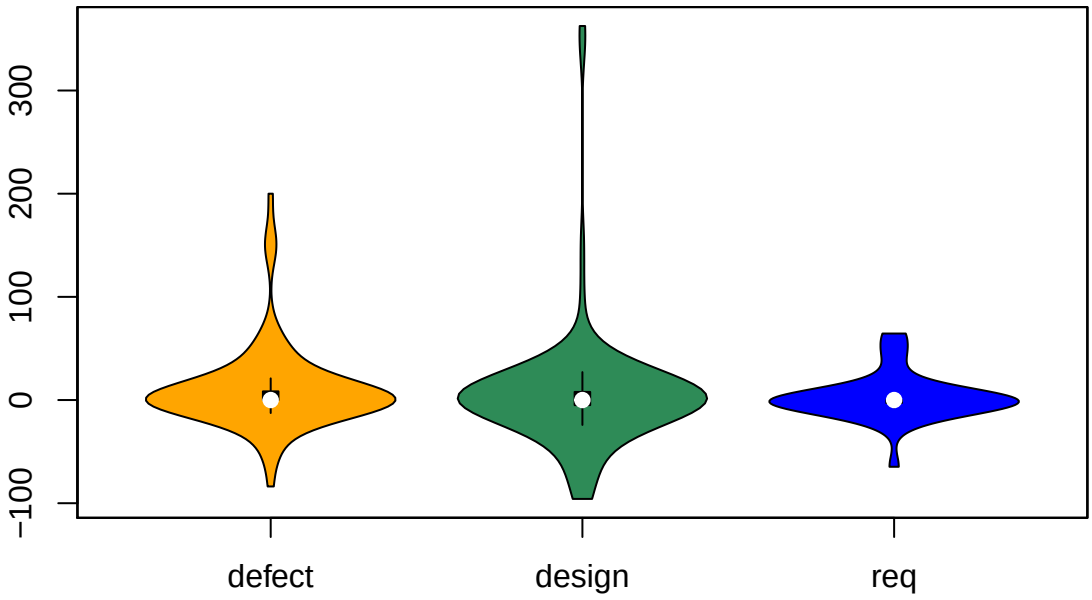
```
## Loading required package: sm
```

```
## Package 'sm', version 2.2-5.4: type help(sm) for summary information
```

```
plot(0, 0, type = "n", xlab = "", ylab = "", axes = FALSE,
     xlim = c(0.5, 3.5), ylim = range(c(defect, design, req)))

axis(side = 1, at = 1:3, labels = c("defect", "design", "req"))
axis(side = 2)

vioplot(defect, at = 1, col = "orange", add = TRUE)
vioplot(design, at = 2, col = "seagreen", add = TRUE)
vioplot(req, at = 3, col = "blue", add = TRUE)
```



Observation

- There is no difference in each category.

Stat

```
b <- data.CountLine.positive
tmp <- b[b[,"Project"] == "jruby", ]
fc <- factor(tmp$Type)
interest <- tmp$interest
tapply(interest, fc, summary)
```

```
## $DEFECT
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.25   20.00   36.36   51.95   56.25  200.00
##
## $DESIGN
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.266   5.634  15.190  37.350  30.750 362.500
##
```

```
## $DOCUMENTATION
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    22.5    22.5    22.5    22.5    22.5    22.5
##
## $REQUIREMENT
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.493   6.849  23.730  30.520  45.450  64.440
```

```
b <- data.CountInput.positive
tmp <- b[b[,"Project"] == "jruby", ]
fc <- factor(tmp$Type)
interest <- tmp$interest
tapply(interest, fc, summary)
```

```
## $DEFECT
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.11   25.00   40.00   96.97  111.90  400.00
##
## $DESIGN
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.143  25.000  36.360  84.460 100.000 400.000
##
## $DOCUMENTATION
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.29   14.29   14.29   14.29   14.29   14.29
##
## $REQUIREMENT
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.143   8.013  20.000  93.260  91.670 450.000
```

Plot

```
library(reshape2)
library(ggplot2)

if(0){
b <- data.CountLine.positive

tmp <- b[b[,"Project"] == "jruby", ]
tmp <- tmp[(tmp[,"Type"] == "DEFECT"),]
a1 <- data.frame(Interest=tmp$interest)
g = ggplot(a1, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.03) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq2-defect.pdf", plot = g, width = 8.09, height = 5)

tmp <- b[b[,"Project"] == "jruby", ]
tmp <- tmp[(tmp[,"Type"] == "DESIGN"),]
a2 <- data.frame(Interest=tmp$interest)
g = ggplot(a2, aes(x=Interest, y=..density.., fill=T, lims(x = c(0,400))))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.03) + guides(fill=FALSE)
```

```
print(g)
ggsave(file = "./tex/figures/rq2-design.pdf", plot = g, width = 8.09, height = 5)

tmp <- b[b,"Project"] == "jruby", ]
tmp <- tmp[(tmp[,"Type"] == "REQUIREMENT"),]
g = ggplot(a3, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.03) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq2-requirement.pdf", plot = g, width = 8.09, height = 5)

b <- data.CountInput.positive

tmp <- b[b[,"Project"] == "jruby", ]
tmp <- tmp[(tmp[,"Type"] == "DEFECT"),]
a1 <- data.frame(Interest=tmp$interest)
g = ggplot(a1, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.03) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq2-defect-countinput.pdf", plot = g, width = 8.09, height = 5)

tmp <- b[b[,"Project"] == "jruby", ]
tmp <- tmp[(tmp[,"Type"] == "DESIGN"),]
a2 <- data.frame(Interest=tmp$interest)
g = ggplot(a2, aes(x=Interest, y=..density.., fill=T, lims(x = c(0,400))))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.03) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq2-design-countinput.pdf", plot = g, width = 8.09, height = 5)

tmp <- b[b[,"Project"] == "jruby", ]
tmp <- tmp[(tmp[,"Type"] == "REQUIREMENT"),]
g = ggplot(a3, aes(x=Interest, y=..density.., fill=T), lims(x = c(0,400)))
g = g + geom_density(alpha = 0.5) + xlim(0, 400) + ylim(0,0.03) + guides(fill=FALSE)
print(g)
ggsave(file = "./tex/figures/rq2-requirement-countinput.pdf", plot = g, width = 8.09, height = 5)
}
```