

2-exploration.R

oomeir

Mon Dec 4 12:23:55 2017

```
library(ggplot2)
library(ggthemes)
library(dplyr)

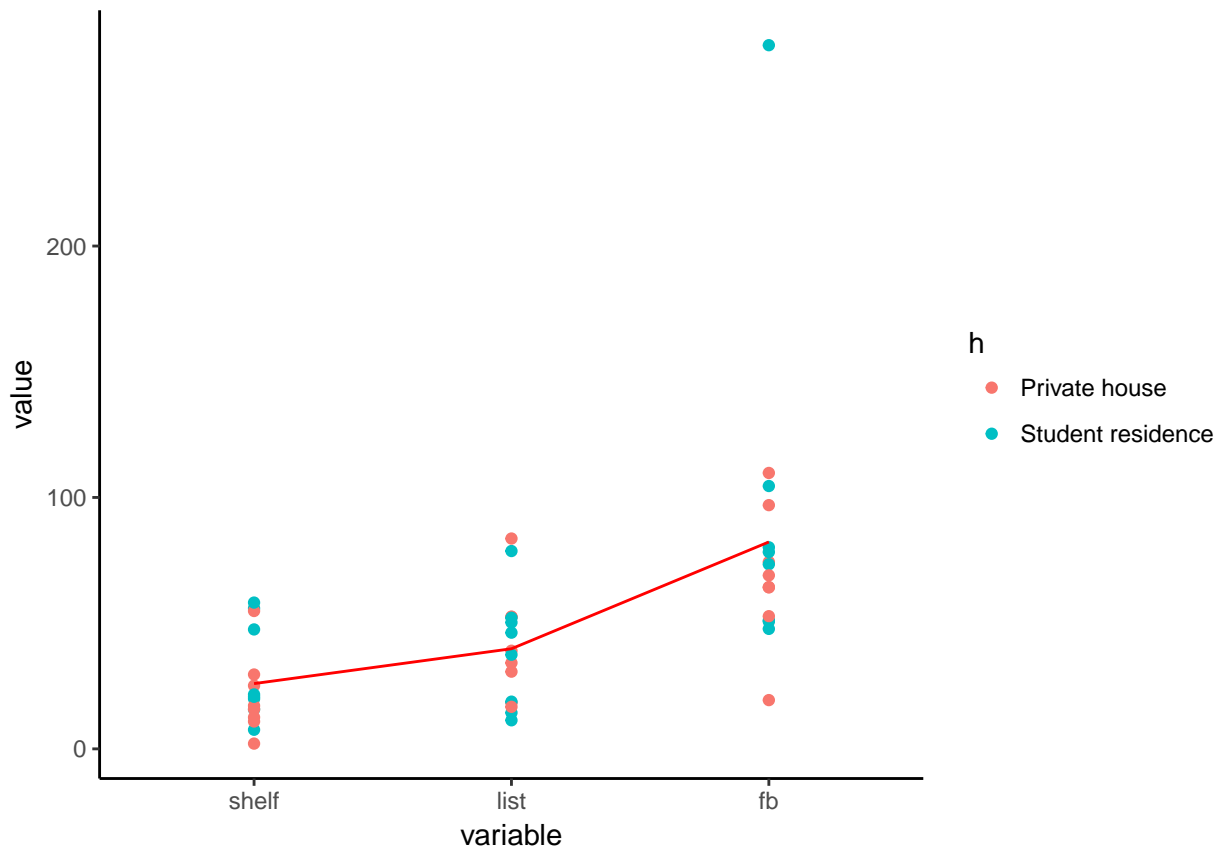
## Warning: package 'dplyr' was built under R version 3.4.1
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
library(stringr)
library(reshape2)
library(knitr)
library(xtable)
library(lubridate)

## Warning: package 'lubridate' was built under R version 3.4.2
##
## Attaching package: 'lubridate'
##
## The following object is masked from 'package:base':
##
##   date
dat <- read.csv("data/data-ready.text", sep = "|", check.names = F)

# Split quant. data into groups for time, satisfaction, and preference
dat.time <- dat[c("participant-number", "house-type", "t1", "t2", "t3", "t4", "t5")]
names(dat.time) <- c("p", "h", "shelf", "list", "fb", "m_return", "fb_return")
dat.satisfaction <- dat[c("participant-number", "t1q", "t2q", "t3q", "t4q", "t5q")]
names(dat.satisfaction) <- c("p", "shelf", "list", "fb", "m_return", "fb_return")
dat.preference <- dat[c("participant-number", "q1", "q2", "q3", "q4")]
names(dat.preference) <- c("p", "machine_usefulness", "fb_usefulness", "shelf_vs_list", "app_vs_fb")

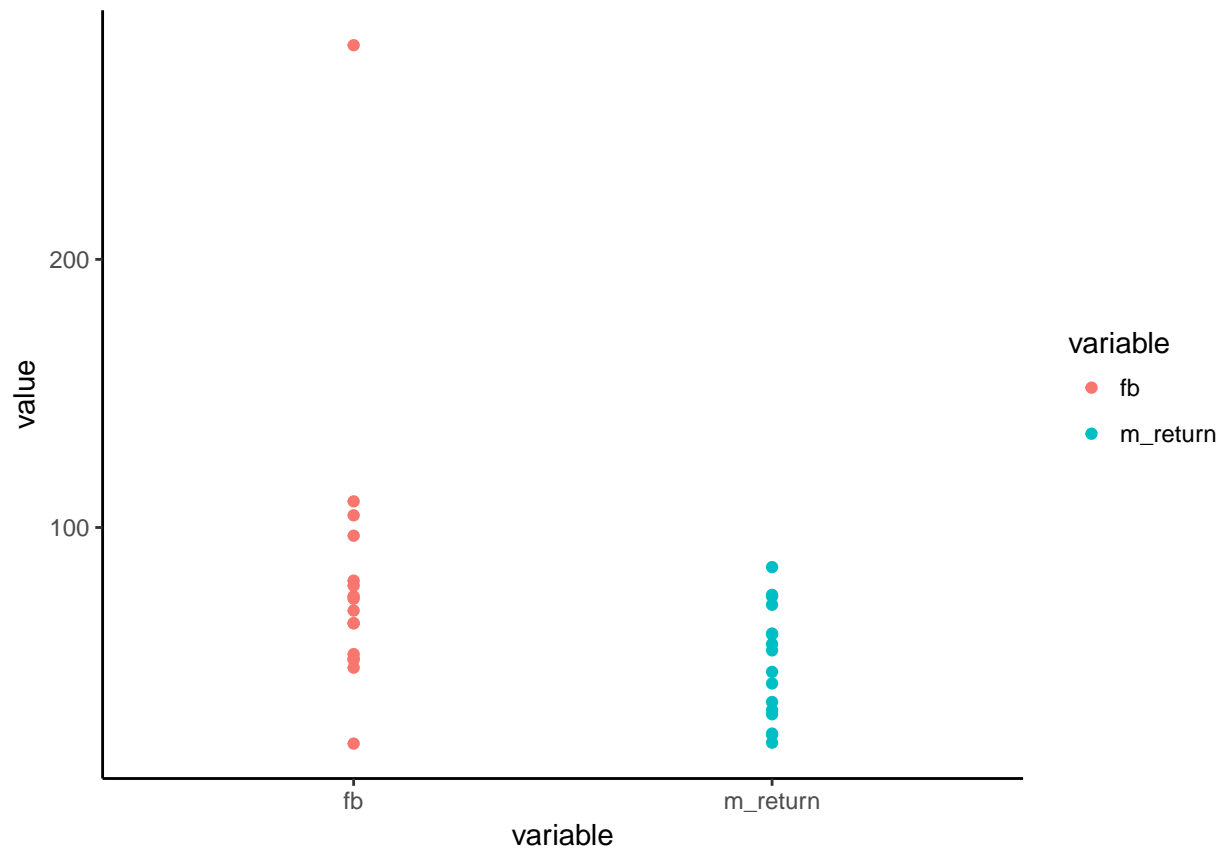
# Plots for most important measures
ggplot(melt(dat.time[2:5]), aes(x=variable, y=value, color=h)) +
  geom_point() +
  stat_summary(aes(y = value, group=1), fun.y=mean, colour="red", geom="line", group=1)+
  theme_classic()

## Using h as id variables
```



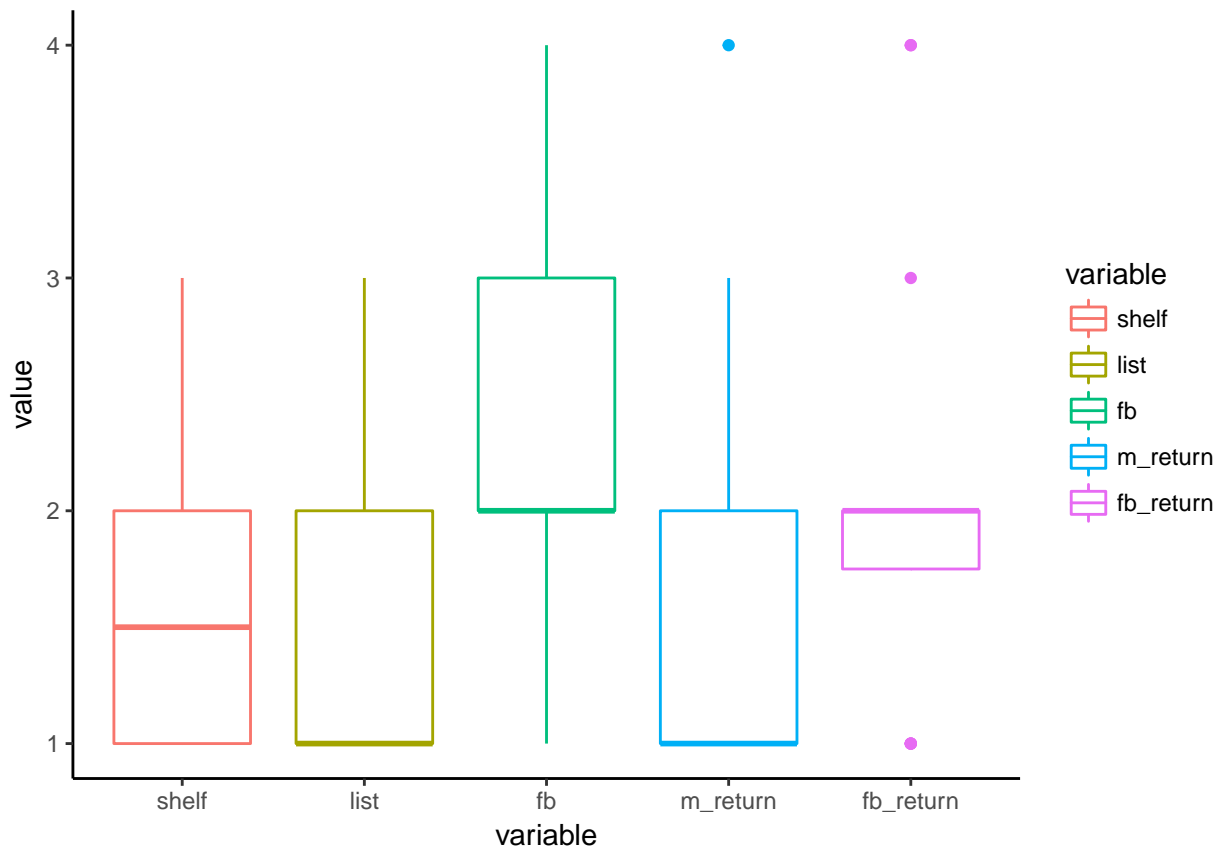
```
ggplot(melt(dat.time[5:6]),aes(x=variable,y=value, color=variable)) +
  geom_point() +
  theme_classic()
```

```
## No id variables; using all as measure variables
```



```
ggplot(melt(dat.satisfaction[2:6]),aes(x=variable,y=value, color=variable)) +
  geom_boxplot() +
  theme_classic()
```

```
## No id variables; using all as measure variables
```



```
# summary stats
mean(dat.satisfaction$shelf)

## [1] 1.625

mean(dat.satisfaction$list)

## [1] 1.375

mean(dat.satisfaction$fb)

## [1] 2.1875

mean(dat.satisfaction$fb_return)

## [1] 2.0625

mean(dat.satisfaction$m_return)

## [1] 1.625

median(dat.satisfaction$shelf)

## [1] 1.5

median(dat.satisfaction$list)

## [1] 1

median(dat.satisfaction$fb)

## [1] 2
```

```

median(dat.satisfaction$fb_return)

## [1] 2
median(dat.satisfaction$m_return)

## [1] 1
range(dat.satisfaction$shelf)

## [1] 1 3
range(dat.satisfaction$list)

## [1] 1 3
range(dat.satisfaction$fb)

## [1] 1 4
range(dat.satisfaction$fb_return)

## [1] 1 4
range(dat.satisfaction$m_return)

## [1] 1 4
# Some tests
t.test(dat.time$shelf, dat.time$list)

##
## Welch Two Sample t-test
##
## data: dat.time$shelf and dat.time$list
## t = -1.96, df = 29.073, p-value = 0.05964
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -28.3227427 0.6009927
## sample estimates:
## mean of x mean of y
## 25.95362 39.81450

# chisq.test(dat.satisfaction$fb, dat.satisfaction$list, correct = FALSE)

# wilcox.test(dat.satisfaction$list, dat.satisfaction$fb)

# Prepare for ANOVA
dat.aov <- melt(dat.time[2:5])

## Using h as id variables
fit <- lm(formula = value ~ variable * h, data = dat.aov)
anova(fit)

## Analysis of Variance Table
##
## Response: value
##


|          | Df | Sum Sq | Mean Sq | F value | Pr(>F)        |
|----------|----|--------|---------|---------|---------------|
| variable | 2  | 27575  | 13787.4 | 9.9525  | 0.0002897 *** |
| h        | 1  | 1584   | 1584.3  | 1.1436  | 0.2910005     |


```

```

## variable:h 2 1719 859.6 0.6205 0.5425175
## Residuals 42 58183 1385.3
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

aov.fit <- aov(formula = value ~ variable * h, data = dat.aov)
summary(aov.fit)

##           Df Sum Sq Mean Sq F value Pr(>F)
## variable    2 27575 13787 9.952 0.00029 ***
## h            1 1584 1584 1.144 0.29100
## variable:h    2 1719 860 0.621 0.54252
## Residuals   42 58183 1385
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TukeyHSD(aov.fit, "variable", ordered = TRUE)

## Tukey multiple comparisons of means
## 95% family-wise confidence level
## factor levels have been ordered
##
## Fit: aov(formula = value ~ variable * h, data = dat.aov)
##
## $variable
##           diff          lwr          upr          p adj
## list-shelf 13.86087 -18.10937 45.83112 0.5479598
## fb-shelf   56.33731 24.36707 88.30755 0.0003055
## fb-list    42.47644 10.50620 74.44668 0.0067032

# Significant effect of interface, but no interaction effect of housing type and interface

# Prepare data to run friedman test of satisfaction
dat.fr <- as.matrix(dat.satisfaction[2:4])
friedman.test(dat.fr)

##
## Friedman rank sum test
##
## data: dat.fr
## Friedman chi-squared = 9.2558, df = 2, p-value = 0.009775

wilcox.test(dat.satisfaction$list, dat.satisfaction$fb)

## Warning in wilcox.test.default(dat.satisfaction$list, dat.satisfaction$fb):
## cannot compute exact p-value with ties
##
## Wilcoxon rank sum test with continuity correction
##
## data: dat.satisfaction$list and dat.satisfaction$fb
## W = 57.5, p-value = 0.004455
## alternative hypothesis: true location shift is not equal to 0

wilcox.test(dat.satisfaction$shelf, dat.satisfaction$fb)

## Warning in wilcox.test.default(dat.satisfaction$shelf, dat.satisfaction
## $fb): cannot compute exact p-value with ties

```

```

##
## Wilcoxon rank sum test with continuity correction
##
## data: dat.satisfaction$shelf and dat.satisfaction$fb
## W = 80, p-value = 0.05493
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(dat.satisfaction$list, dat.satisfaction$shelf)

## Warning in wilcox.test.default(dat.satisfaction$list, dat.satisfaction
## $shelf): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: dat.satisfaction$list and dat.satisfaction$shelf
## W = 103, p-value = 0.2895
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(dat.preference$machine_usefulness, dat.preference$fb_usefulness)

## Warning in wilcox.test.default(dat.preference$machine_usefulness,
## dat.preference$fb_usefulness): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: dat.preference$machine_usefulness and dat.preference$fb_usefulness
## W = 78, p-value = 0.03394
## alternative hypothesis: true location shift is not equal to 0

```