

TRANSLATOR Opposites

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
data <- read.table("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/experimental-outputs/")
median.time.in.seconds <- median(data$V1)
median.time.in.seconds
```

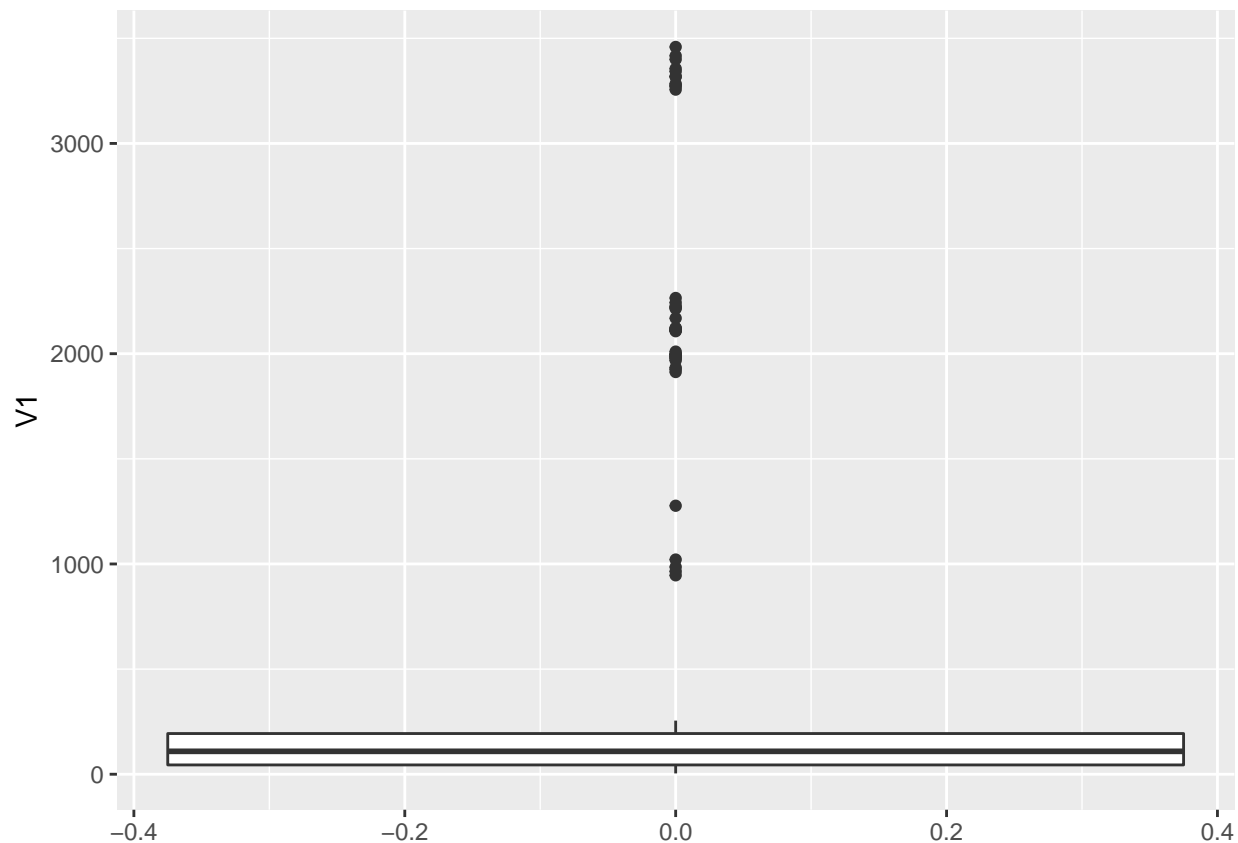
```
## [1] 109
```

```
median.tasks.per.minute <- 60 / median.time.in.seconds
pay.per.task <- 0.10
median.pay.per.hour <- 60 * (median.tasks.per.minute * pay.per.task)

print(paste("At 60 seconds per minute, a median time of ", median.time.in.seconds, "seconds per task yields a median of ", median.pay.per.hour, "dollars per hour"))
```

```
## [1] "At 60 seconds per minute, a median time of 109 seconds per task yields a median of 0.550458715 dollars per hour"
```

Across the entire set of Turkers, the variability in time per assignment was enormous, and the median was quite high. Consequently, the median hourly rate was quite low.



Breaking the time per assignment down by annotator, we see that most of the annotators were actually fast. However, one slow annotator accounted for most of the HITs. Consequently, most annotators received an hourly rate that is *much* higher than the median.

Oh, wait, that's not true. In fact most annotators received a low hourly rate of a bit over \$3/hour, while one annotator got a *terribly* low rate, and unfortunately did most of the HITs.

```
# the spreadsheet is hairy, so use fread with fill = TRUE
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.0.2
```

```
amt <- fread("/Users/kevincohen/Dropbox/N-Z/translator-concept-oppositeness/experimental-outputs/Batch 1")
#head(amt)
amt$WorkerId <- as.factor(amt$WorkerId)
amt$Input.word01 <- as.factor(amt$Input.word01)
amt$Input.word02 <- as.factor(amt$Input.word02)
amt$Input.goldResponse <- as.factor(amt$Input.goldResponse)
amt$Input.goldVerify <- as.factor(amt$Input.goldVerify)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.2
```

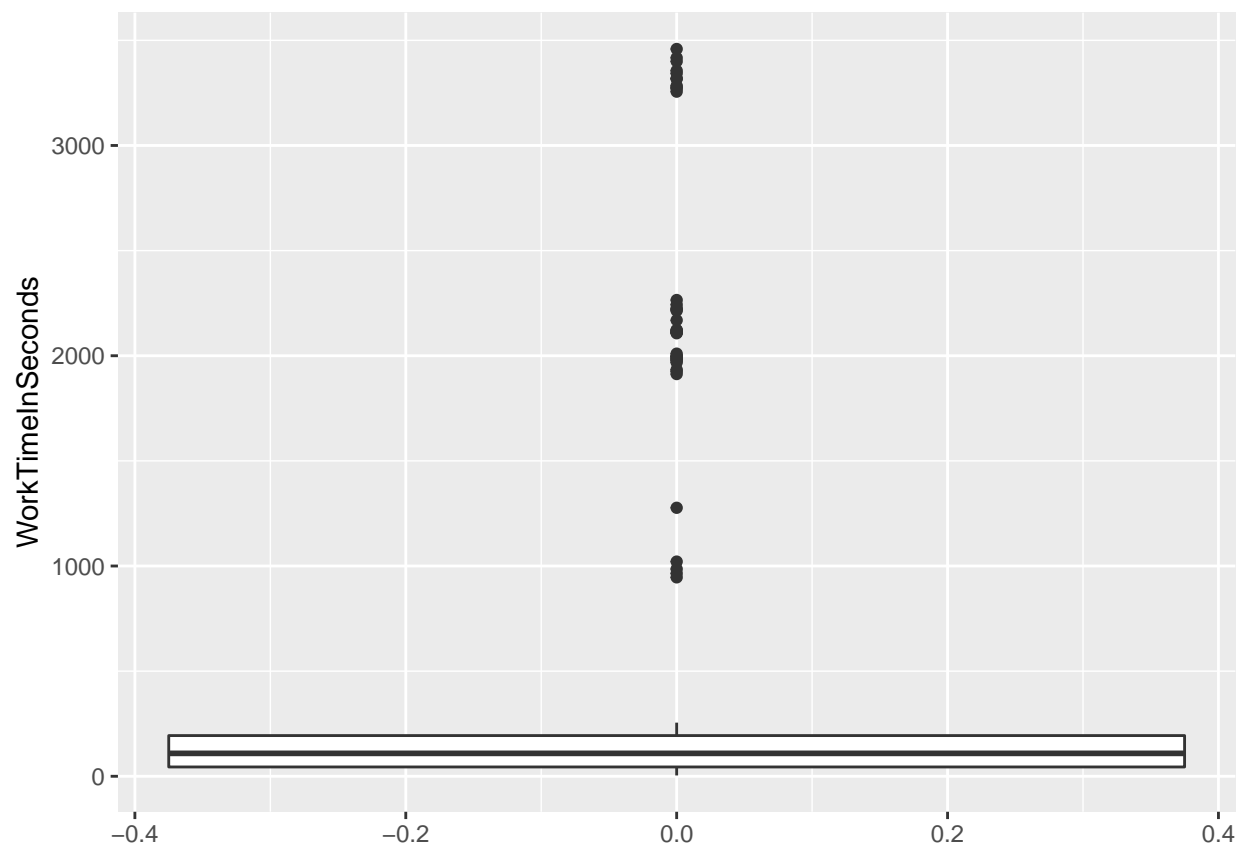
```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##   between, first, last

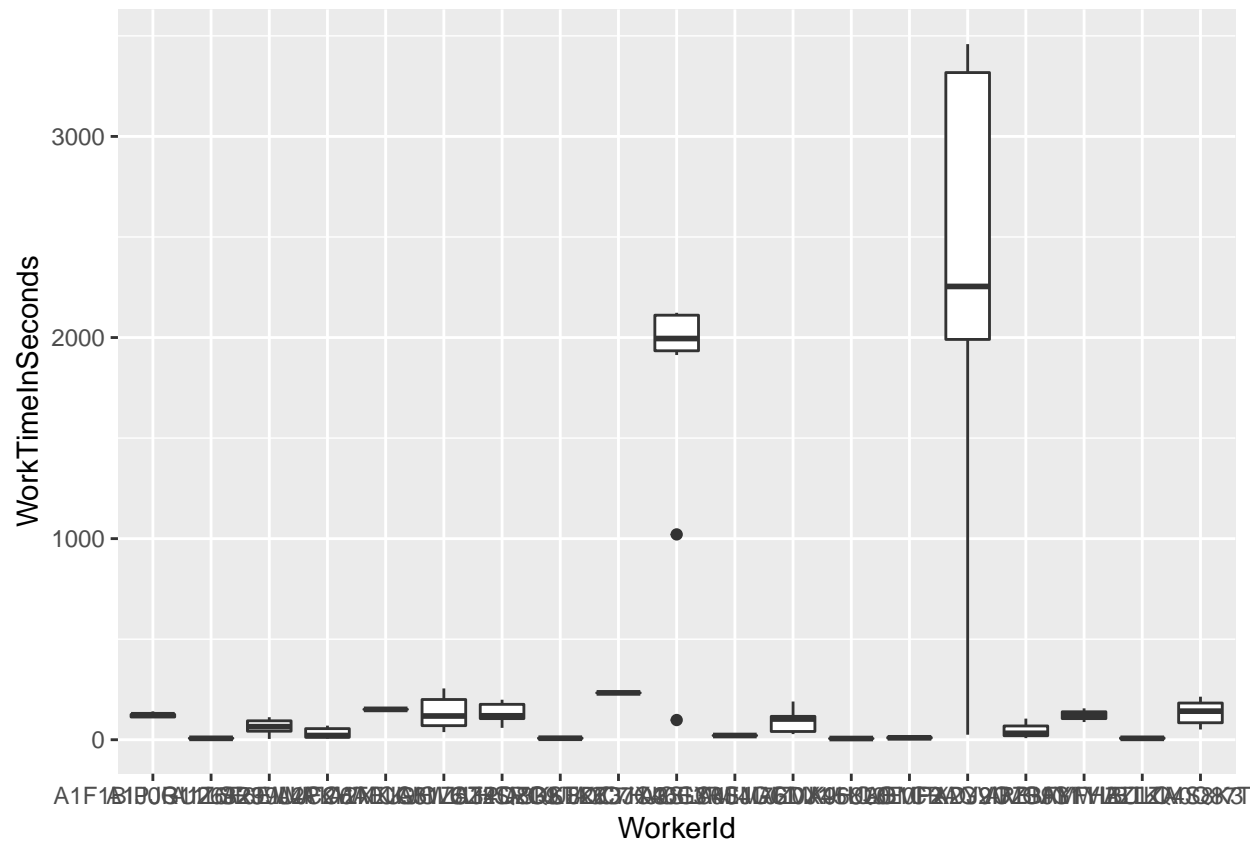
## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
by.turker <- amt %>% group_by(WorkerId)
#head(by.turker)
ggplot(amt, aes(y = WorkTimeInSeconds)) + geom_boxplot()
```



```
ggplot(amt, aes(x = WorkerId, y = WorkTimeInSeconds)) + geom_boxplot()
```



does this look any different if I plot the grouped

```
ggplot(by.turker, aes(x = WorkerId, y = WorkTimeInSeconds)) + geom_boxplot()
```

