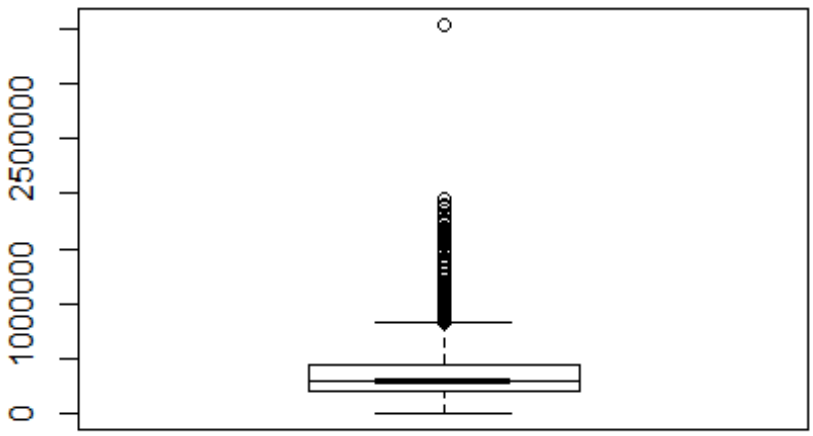
# Data Exploration Project

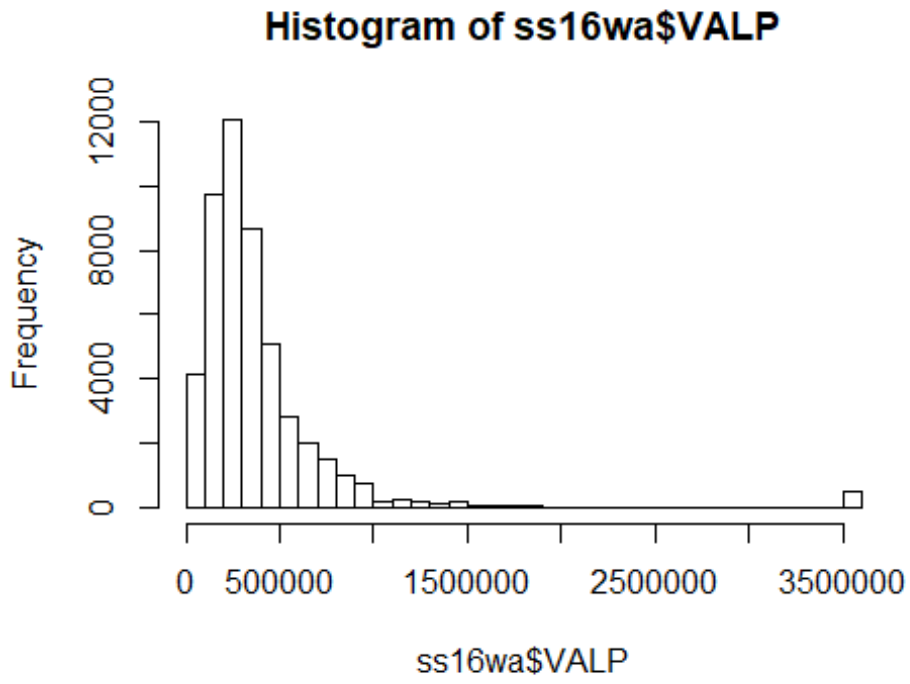
Lukas Fontanilla

Tue Jun 12 15:29:07 2018

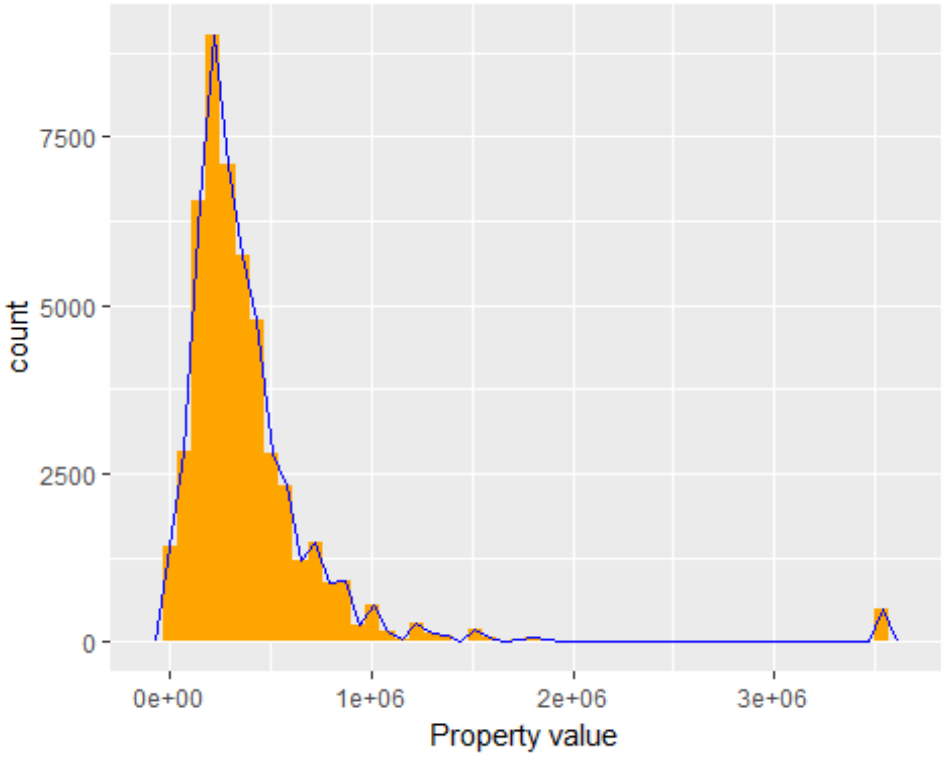
```
# read in file
ss16wa <- read.csv("ss16wa.csv")

#print summary column of "VALP"
summary("VALP")

##    Length     Class      Mode
##         1 character character

#draw box-and-whisker plot with a notch for "VALP"
boxplot(ss16wa$VALP, notch = TRUE)
```



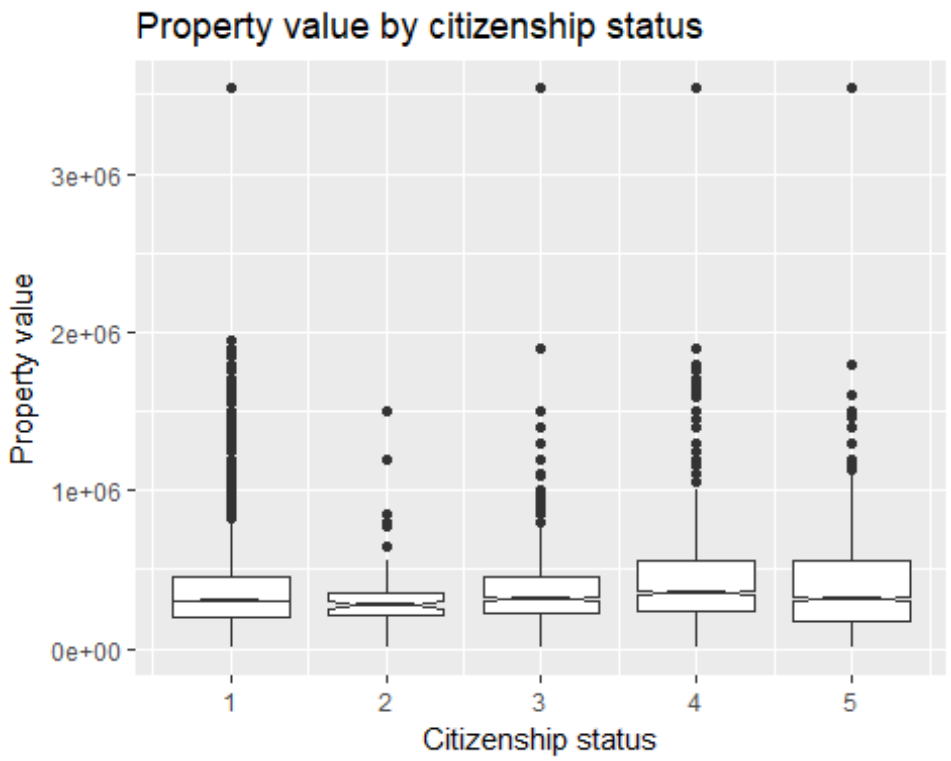

```
#draw histogram with 50 bins
hist(ss16wa$VALP,breaks = 50)

# draw a histogram and frequency polygon for "VALP"
# 50 bins for both, VALP values x-axis (lable: "Property value")
# hist bars orange, polygon line blue
library(ggplot2)
```




```
ggplot(ss16wa, aes(VALP)) +
  geom_histogram(mapping = NULL, data = NULL, stat = "bin", bins = 50, fill =
 "orange") +
  geom_freqpoly(mapping = NULL, data = NULL, stat = "bin", bins = 50, colour
= "blue") +
  xlab("Property value")
```




```
# convert CIT (citizenship status) to factor
CIT_convert <- as.factor(ss16wa$CIT)

# box and whiskers plots for valp by cit
ggplot(ss16wa, aes(CIT, VALP, group = CIT)) +
  geom_boxplot(mapping = NULL, data = NULL, stat = "boxplot",
               notch = TRUE) +
  labs(x = "Citizenship status", y = "Property value") +
  ggtitle("Property value by citizenship status")
```


Property value by citizenship status
Property value
3e+06
2e+06
1e+06
0e+00
1
2
3
4
5
Citizenship status