Data Exploration in R

Meredith Dodd <meredith.dodd@woz-u.com>

Tue 4/28/2020 4:49 PM

To: Meredith Dodd <meredith.dodd@woz-u.com>

```
# Data Exploration

library("ggplot2")
library("dplyr")

#Find the different artists in the dataset - exploring categorical, i.e. factor data
levels(top10s$artist)

#Find all the years in this particular dataset

unique(top10s$year)

head(top10s)

tail(top10s)

#Boxplot of beats per minute

ggplot(top10s, aes(x =artist, y=bpm)) + geom_boxplot()

#That's way too crazy! Let's filter and only look at a few artists! Choose five!

Ladies <- top10s %>% filter(artist %in% c("Lady Gaga", "Rihanna", "Alicia Keys", "Katy Perry", "Britney Spears"))

ggplot(Ladies, aes(x =artist, y=bpm)) + geom_boxplot()

#How about you graph with artist as a color?

ggplot(Ladies, aes(x =year, y=bpm, color = artist)) + geom_line()


#Highest bpm in 2010? 2019?

filter(Ladies, year %in% c(2010,2019)) %>% arrange(desc(bpm))

#What if you don't wnat all that noise? Well, you can select columns too

filter(Ladies, year %in% c(2010,2019)) %>% arrange(desc(bpm)) %>% select(artist,year,bpm)

#Get the median beats per minute

median <- Ladies %>% select(year, bpm) %>% group_by(year) %>% summarise(medBPM = median(bpm))
```

Meredith Dodd, Ph.D. | Data Science Program Chair and Instructor
meredith.dodd@woz-u.com
o: 480-291-8068

powered by WOZ U

https://woz-u.com