HANDS-ON R

October 4, 2018

James Ng, PhD
james.ng@nd.edu
Tutorials

1. Exploring World Cup data
2. Plotting graphs using ggplot2
3. Factors
4. Repetition using for loop, apply family, function
TUTORIAL 1
Exploring World Cup data
install.packages('faraway')

library(faraway)

dat <- data.frame(worldcup)

head(dat)
str(dat)
summary(dat)
1. Retrieve the value in row 17, column 3.

2. Retrieve the first five columns for the first six rows.

3. Retrieve values by row and column names.

4. Retrieve all column values for row Alonso.

5. Retrieve all row values for column Team.

6. Extract the row names and store them in a new column Player.

7. Reorder column Player to the furthest left.

8. What's the max number of shots taken on each team?

9. Which player took the most shots on each team?
TUTORIAL 2
Plotting graphs with ggplot2
1. Start a new R Script.
2. Clear objects in environment, set working directory
   ```r
   rm(list=ls())
   setwd(‘/Users/jng2/Dropbox/Work/Library/CDS/R-RStudio/hands-on’)
   ```
3. Load ‘tidyverse’ (may have to install first)
   ```r
   library(tidyverse)
   ```
4. Reload ‘worldcup’ data
TUTORIAL 2: plotting (cont’d)

Total passes vs total shots
Number of shots by position

# A tibble: 4 x 5
  Position passes shots tackles saves
    <fct>    <int>   <int> <int>
1 Defender  19297   219    1027     0
2 Forward   7268   605     289     0
3 Goalkeeper 2003     1     1   397
4 Midfielder 21722  546   1177     0
TUTORIAL 2: plotting (cont’d)

Moves (shots, passes, tackles, saves) by position

<table>
<thead>
<tr>
<th>Position</th>
<th>move</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defender</td>
<td>passes</td>
<td>19297</td>
</tr>
<tr>
<td></td>
<td>shots</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td>tackles</td>
<td>1027</td>
</tr>
<tr>
<td></td>
<td>saves</td>
<td>0</td>
</tr>
<tr>
<td>Forward</td>
<td>passes</td>
<td>7268</td>
</tr>
<tr>
<td></td>
<td>shots</td>
<td>605</td>
</tr>
<tr>
<td></td>
<td>tackles</td>
<td>289</td>
</tr>
<tr>
<td></td>
<td>saves</td>
<td>0</td>
</tr>
<tr>
<td>Goalkeeper</td>
<td>passes</td>
<td>2003</td>
</tr>
<tr>
<td></td>
<td>shots</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>tackles</td>
<td>1177</td>
</tr>
<tr>
<td></td>
<td>saves</td>
<td>397</td>
</tr>
<tr>
<td>Midfielder</td>
<td>passes</td>
<td>21722</td>
</tr>
<tr>
<td></td>
<td>shots</td>
<td>546</td>
</tr>
<tr>
<td></td>
<td>tackles</td>
<td>1027</td>
</tr>
<tr>
<td></td>
<td>saves</td>
<td>0</td>
</tr>
</tbody>
</table>
Moves by position: faceting
TUTORIAL 2: plotting (cont’d)

Moves (%) by position

<table>
<thead>
<tr>
<th>Position</th>
<th>move</th>
<th>count</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defender</td>
<td>passes</td>
<td>19297</td>
<td>0.939</td>
</tr>
<tr>
<td>Forward</td>
<td>passes</td>
<td>7268</td>
<td>0.890</td>
</tr>
<tr>
<td>Goalkeeper</td>
<td>passes</td>
<td>2003</td>
<td>0.834</td>
</tr>
<tr>
<td>Midfielder</td>
<td>passes</td>
<td>21722</td>
<td>0.927</td>
</tr>
<tr>
<td>Defender</td>
<td>shots</td>
<td>219</td>
<td>0.0107</td>
</tr>
<tr>
<td>Forward</td>
<td>shots</td>
<td>605</td>
<td>0.0741</td>
</tr>
<tr>
<td>Goalkeeper</td>
<td>shots</td>
<td>1</td>
<td>0.000416</td>
</tr>
<tr>
<td>Midfielder</td>
<td>shots</td>
<td>546</td>
<td>0.0233</td>
</tr>
<tr>
<td>Defender</td>
<td>tackles</td>
<td>1027</td>
<td>0.0500</td>
</tr>
<tr>
<td>Forward</td>
<td>tackles</td>
<td>289</td>
<td>0.0354</td>
</tr>
<tr>
<td>Goalkeeper</td>
<td>tackles</td>
<td>1177</td>
<td>0.0502</td>
</tr>
<tr>
<td>Midfielder</td>
<td>tackles</td>
<td>397</td>
<td>0.165</td>
</tr>
<tr>
<td>Defender</td>
<td>saves</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Forward</td>
<td>saves</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Goalkeeper</td>
<td>saves</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Midfielder</td>
<td>saves</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
TUTORIAL 2: plotting (cont’d)

Moves (%) by position

# first compute percentage of each move taken per position

```r
datplot3b <- datplot3 %>%
  group_by(Position) %>%
  mutate(percent = count/sum(count))
```

# then plot the graph

```r
ggplot(datplot3b, aes(x=Position, y=percent, fill=move)) +
  geom_bar(stat="identity", position="fill")
```
Positions (%) by move

datplot3c <- datplot3 %>%
  group_by(move) %>%
  mutate(percent = count/sum(count))

ggplot(datplot3c, aes(x=move, y=percent, fill=Position)) +
geom_bar(stat="identity", position="fill")
TUTORIAL 3
Factors
1. Continuing using ‘worldcup’ data.

2. Team and Position are factor variables – R knows how to deal with factors properly

summary(dat$Team)
summary(dat$Position)
levels(dat$Position)
table(dat$Team, dat$Position)

mod <- lm(Shots ~ Position + Time + Passes + Tackles + Saves, data=dat)
summary(mod)
1. Sometimes factor levels should be ordered
2. Create a factor variable categorizing amount of shooting: low, medium, high, extremely high
3. Order the levels of this factor variable such that low < medium < high < extremely high
TUTORIAL 4
Repetition using for loop, apply family, function
1. Load ‘iris’ data

2. Compute the number of unique values in each column in iris using
   - For loop
   - apply
   - lapply
   - sapply

This is what you should obtain:

<table>
<thead>
<tr>
<th>Sepal.Length</th>
<th>Sepal.Width</th>
<th>Petal.Length</th>
<th>Petal.Width</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>23</td>
<td>43</td>
<td>22</td>
<td>3</td>
</tr>
</tbody>
</table>