

DA Exercises

Hira Piracha

##Exercise 1.1

```
getwd()
```

```
## [1] "/Users/hirapiracha/Desktop"
```

```
hotels_vienna <- read.csv(file="/Users/hirapiracha/R-Coding1/hotels-vienna.csv", header=TRUE, sep=",")
```

1.)

```
Test1 <- hotels_vienna[sample(nrow(hotels_vienna), 25), ]  
(Average_1 = mean(Test1$price, na.rm = TRUE))
```

```
## [1] 147.56
```

2.

```
Test2 <- hotels_vienna[sample(nrow(hotels_vienna), 50), ]  
(Average_2 = mean(Test2$price, na.rm = TRUE))
```

```
## [1] 126.34
```

3.

```
Test3 <- hotels_vienna[sample(nrow(hotels_vienna), 200), ]  
(Average_3 = mean(Test3$price, na.rm = TRUE))
```

```
## [1] 136.06
```

4.

```
(Average_4 = mean(hotels_vienna$price, na.rm = TRUE))
```

```
## [1] 131.3668
```

#Summary: The outcomes were different each time I ran the sample. #There is a price difference between the average hotel prices of the data and the sampling. However, more testing is needed for further analysis.

##Exercise 2.5

```
library(dplyr)  
ep1_games <- read.csv(file="/Users/hirapiracha/R-Coding1/ep1_games.csv", header=TRUE, sep=",")  
head(ep1_games)
```

##	div	season	date	team_home	team_away	points_home	points_away
## 1	E0	2008	16aug2008	Arsenal	West Brom	3	0
## 2	E0	2008	16aug2008	West Ham	Wigan	3	0
## 3	E0	2008	16aug2008	Middlesbrough	Tottenham	3	0
## 4	E0	2008	16aug2008	Everton	Blackburn	0	3
## 5	E0	2008	16aug2008	Bolton	Stoke	3	0
## 6	E0	2008	16aug2008	Sunderland	Liverpool	0	3

##	goals_home	goals_away
## 1	1	0
## 2	2	1
## 3	2	1
## 4	2	3
## 5	3	1
## 6	0	1

```
cols( div = col_character(), season = col_double(), date = col_character(), team_home = col_character(),
team_away = col_character(), points_home = col_double(), points_away = col_double(), sgoals_home =
col_double(), goals_away = col_double() )
```

```
st2 <- epl_games %>% group_by(team_home) %>%
  filter(n_distinct(season) == 11) %>%
  summarise(team_home, season)
```

#Teams that played in all seasons:

#Arsenal

#Everton

#Chelsea

#Man United

#Liverpool

#Tottenham

#Man City

```
st3 <- epl_games %>% group_by(team_home) %>%
  filter(n_distinct(season) == 1) %>%
  summarise(team_home, season)
```

#Teams that only played once

#Blackpool

#Reading