

HW2_CThomas_26Sept18

Charlene Thomas

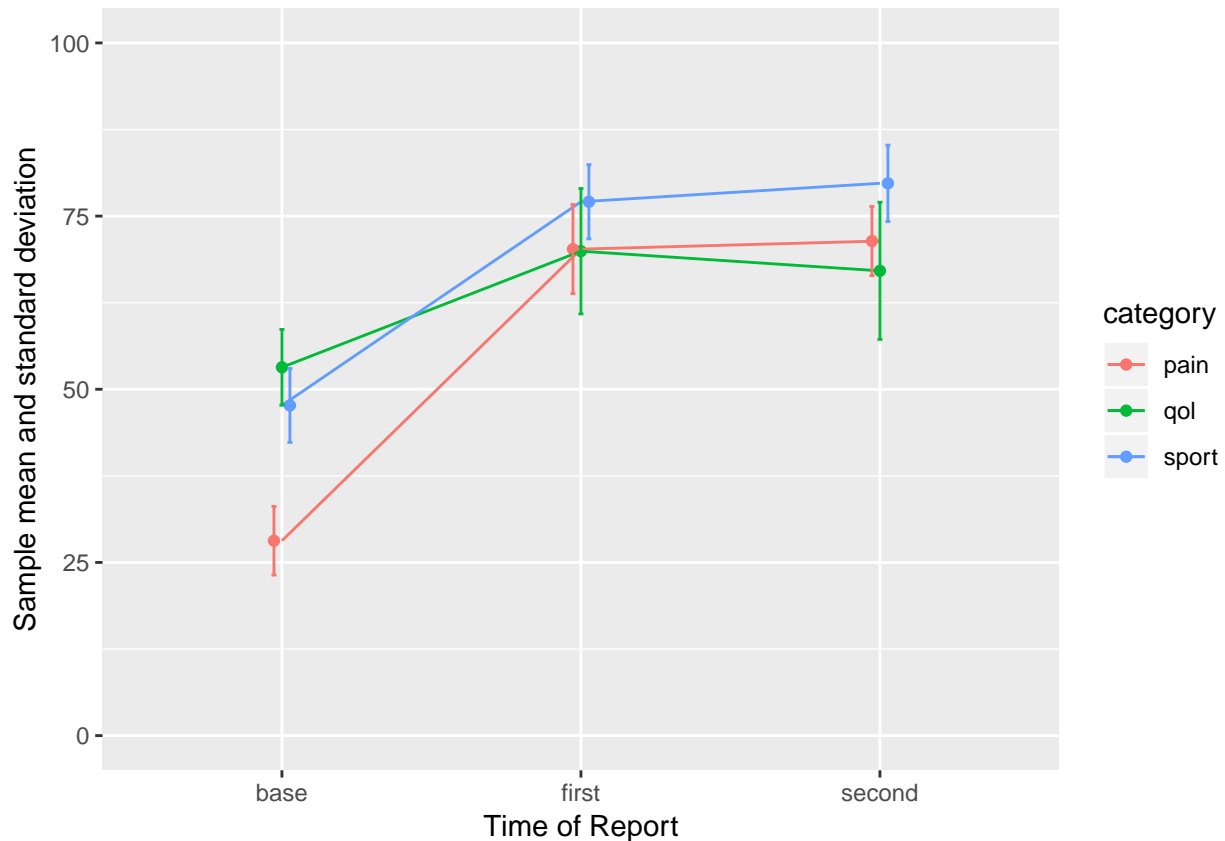
9/26/2018

Assignment for Lab 2

1. Read Chapters 3, 5, 12, 13 from “R for data science” (online version).
2. Create a GitHub repository Lab2 (files with R code for this assignment should be added to this repository, do not push csv files for Problem 4 to the GitHub repo).
3. Write the R code to recreate the graph from the Activity 2 (Problem 4).

```
setwd("~/Documents/BSDS - Fall 2018/Biostatistics I/Data")
activity <- read.csv("lab2.csv")

activity %>%
  gather(`base_sport`, `base_pain`, `base_qol`,
         `first_sport`, `first_pain`, `first_qol`, `second_sport`,
         `second_qol`, `second_pain`, key = "time_category", value = "cases") %>%
  separate(time_category, into = c("time", "category")) %>%
  group_by(time, category) %>%
  summarize(mean = mean(cases), sd = sd(cases)) %>%
  ggplot(aes(x = time, y = mean, color = category, group = category)) +
  geom_point(position = position_dodge(0.08)) +
  geom_line() +
  geom_errorbar(aes(ymin = mean - sd, ymax = mean + sd), width = 0.05,
               position = position_dodge(0.08)) +
  ylim(0,100) +
  labs(x = "Time of Report", y = "Sample mean and standard deviation")
```



4. Download two files.

4.1 Make the data frames “tidy”.

Please refer to R code under 4.2

4.2 Merge two data frames: the resulting data frame should contain information about coverage and expenditures for years 2013-2016. Please note that file expenditures.csv does not contain years 2015-2016.

```
setwd("~/Documents/BSDS - Fall 2018/Biostatistics I/Data")
coverage<- read.csv(file = "coverage.csv", skip =2, check.names = FALSE)
coverage<- coverage[1:52,]

tidy_coverage <- coverage %>%
  gather(`2013__Employer`, `2013__Non-Group`, `2013__Medicaid`, `2013__Medicare`,
    `2013__Other Public`, `2013__Uninsured`, `2013__Total`, `2014__Employer`, `2014__Non-Group`, `2014__Medicaid`,
    `2014__Medicare`, `2014__Other Public`, `2014__Uninsured`, `2014__Total`, `2015__Employer`, `2015__Non-Group`,
    `2015__Medicaid`, `2015__Medicare`, `2015__Other Public`, `2015__Uninsured`, `2015__Total`, `2016__Employer`,
    `2016__Non-Group`, `2016__Medicaid`, `2016__Medicare`, `2016__Other Public`, `2016__Uninsured`,
    key = "year_category", value = "Enrollment (# of People)" %>%
  separate(year_category, into = c("Year", "Coverage Type"))
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

```
## Warning: Too many values at 416 locations: 53, 54, 55, 56, 57, 58, 59, 60,
## 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, ...
```

```
expenditures<- read.csv(file = "expenditures.csv", skip =2,check.names = FALSE)
expenditures<- expenditures[1:52,]
```

```
tidy_expenditures <- expenditures %>%
  gather(`1991__Total Health Spending`, `1992__Total Health Spending`, `1993__Total Health Spending`, `1994__Total Health Spending`, `1995__Total Health Spending`, `1996__Total Health Spending`, `1997__Total Health Spending`, `1998__Total Health Spending`, `1999__Total Health Spending`, `2000__Total Health Spending`, `2001__Total Health Spending`, `2002__Total Health Spending`, `2003__Total Health Spending`, `2004__Total Health Spending`, `2005__Total Health Spending`, `2006__Total Health Spending`, `2007__Total Health Spending`, `2008__Total Health Spending`, `2009__Total Health Spending`, `2010__Total Health Spending`, `2011__Total Health Spending`, `2012__Total Health Spending`, `2013__Total Health Spending`, `2014__Total Health Spending`, key = "year_category", value = "Total Health Spending") %>%
  separate(year_category, into = c("Year", "Type of Spending")) %>%
  select("Location", "Year", "Total Health Spending")
```

```
## Warning: Too many values at 1248 locations: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
## 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...
```

```
tidy_df <- tidy_coverage %>%
left_join(tidy_expenditures)
```

```
## Joining, by = c("Location", "Year")
```

```
## Warning: Column `Location` joining factors with different levels, coercing
## to character vector
```

```
head(tidy_df)
```

```
##      Location Year Coverage Type Enrollment (# of People)
## 1 United States 2013      Employer      155696900
## 2      Alabama 2013      Employer      2126500
## 3       Alaska 2013      Employer      364900
## 4      Arizona 2013      Employer      2883800
## 5    Arkansas 2013      Employer      1128800
## 6   California 2013      Employer      17747300
##      Total Health Spending
## 1      2435624
## 2      33788
## 3      7684
## 4     41481
## 5     20500
## 6     278168
```

5. Submit a link to the repo “Lab2” via Canvas.

Link: <https://github.com/cht2028/Lab2.git>