PT_project_EDA_0404

Fiona

4/5/2020

```
dat1 <-  read_excel("~/Desktop/For Masanao Class - ROMS Full Data Set - March 19th, 2019 Upload.xlsx",sh
dat2 <-  read_excel("~/Desktop/For Masanao Class - ROMS Full Data Set - March 19th, 2019 Upload.xlsx",sh
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i =
## sheet, : Expecting numeric in Z10703 / R10703C26: got 'NULL'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i =
## sheet, : Expecting numeric in Z16107 / R16107C26: got 'unknown'
```

```
# data cleaning part starts here.
# fix the age
dat2$Age <- floor(dat2$Age)
# fix the typo for outcomes
dat2$Outcome[dat2$Outcome =="Neck DISABILITY INDEX" ] <-  "NECK DISABILITY INDEX"
dat2$Outcome[dat2$Outcome =="neck DISABILITY INDEX" ] <-  "NECK DISABILITY INDEX"
# fix the typo for body regions
dat2$`Body Region`[dat2$`Body Region` == "knee"] <- "Knee"
dat2$`Body Region`[dat2$`Body Region` == "lumbar"] <- "Lumbar"
# remove the duplicated rows and select some columns.
dat <- dat2 %>% distinct() %>% select(`ROMS ID`, `Visit ID`, Payer, `Payer Category`, Age, `Sex (1=male
# fix the date.
dat$`Injury Date` <- as.numeric(dat$`Injury Date`)
```

```
## Warning: NAs introduced by coercion
```

```
dat$`Injury Date` <- as.Date(dat$`Injury Date`-25569)
dat$`Surgery Date` <- as.numeric(dat$`Surgery Date`)
```

```
## Warning: NAs introduced by coercion
```

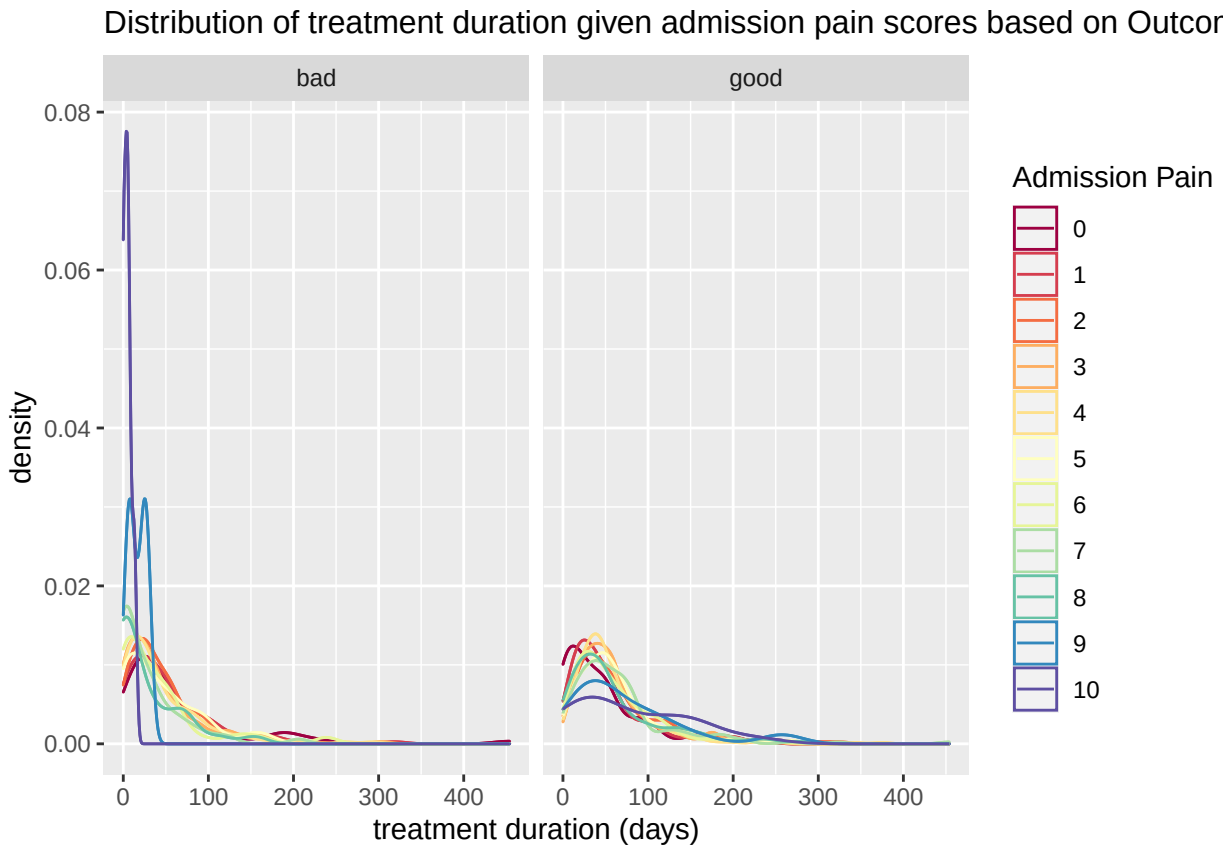
```
dat$`Surgery Date` <- as.Date(dat$`Surgery Date`-25569)
```

```
da<- dat %>% select(`ROMS ID`, Outcome, `Body Region`, Surgical, `Admission Date`,`Discharge Date`, `Adm

dacheck<- dat %>% select(`ROMS ID`, Outcome, `Body Region`, Surgical, `Admission Date`,`Discharge Date`


# evaluate the effectiveness for the pain level.
da_evaluate <- da  %>% mutate(pain_effect = ifelse(`Pain Change Scores`<= -2 & `Admission Pain` >= 2, "g

# evaluate the treatment effectiveness overall.
# check the outcome = "LOWER EXTREMITY FUNC SCALE", 1385 records, filter the wrong scale and get 1341 r
da_evaluate1<- da_evaluate %>%  filter(Outcome == "LOWER EXTREMITY FUNC SCALE") %>% mutate(effect_all =
da_evaluate1 <- da_evaluate1 %>% filter(`Admission Outcome Score`<=80) %>% filter(`Discharge Outcome Sc

# check the outcome = "knee outcome survey" that has 699 records.
da_evaluate2<- da_evaluate %>%  filter(Outcome == "KNEE OUTCOME SURVEY") %>% mutate(effect_all = ifelse
da_evaluate2 <- da_evaluate2 %>% filter(`Admission Outcome Score`<= 100) %>% filter(`Discharge Outcome S
# check the outcome within "MODIFIED LOW BACK DISABILITY QUESTIONNAIRE","Quick DASH","NECK DISABILITY I
da_evaluate3 <- da_evaluate %>%  filter(Outcome %in% c("MODIFIED LOW BACK DISABILITY QUESTIONNAIRE","Qu
```

```
da_evaluate3 <- da_evaluate3 %>% filter(`Admission Outcome Score`<= 100) %>% filter(`Discharge Outcome S
# data after evaluation based on the Outcome's excel formula.
da_eval <- rbind(da_evaluate1,da_evaluate2,da_evaluate3)

# check out if individual went for multiple treatments.
# da1 contains that people who have more than 1 records, like more than one types outcomes, different d
da1 <- da_eval %>% group_by(`ROMS ID`) %>% filter(n() >1)
da2 <- da_eval %>% group_by(`ROMS ID`) %>% filter(n() ==1)
```

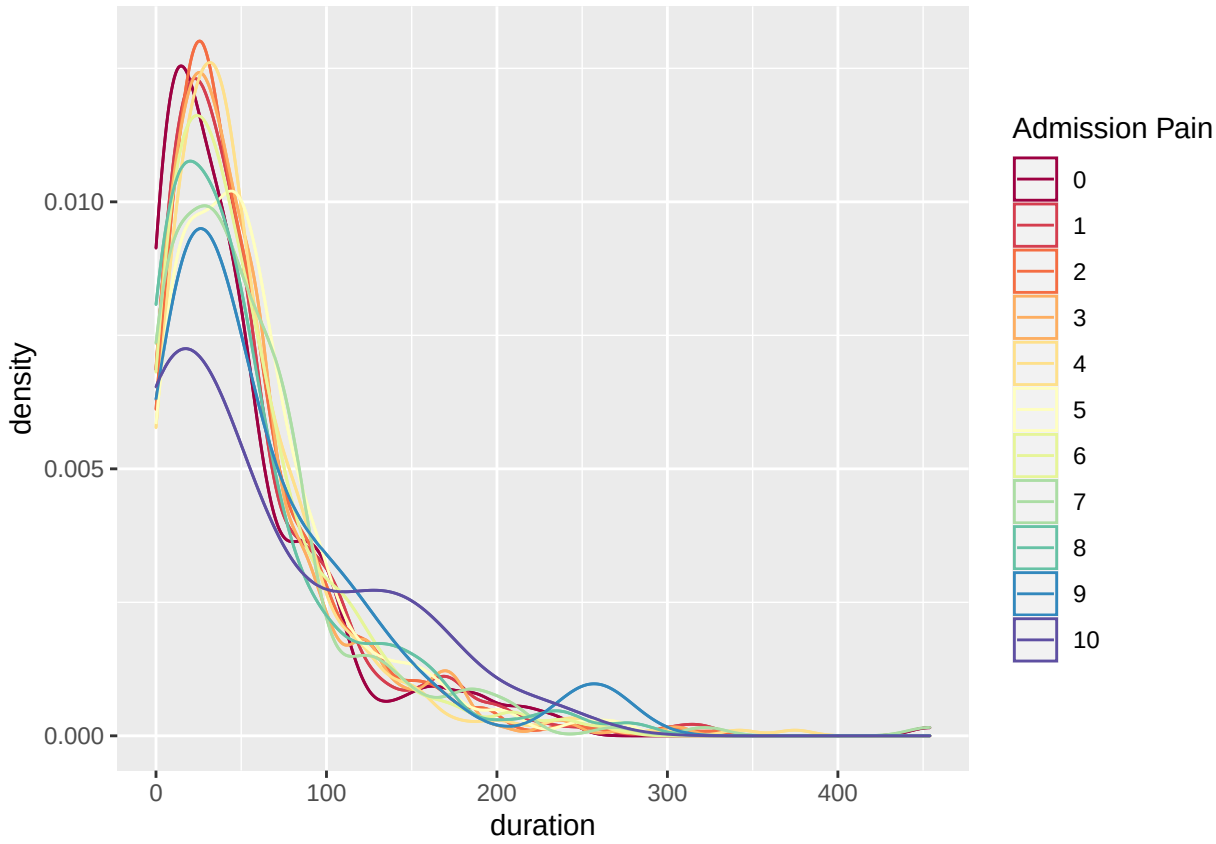
```
# the plot displays the association between the duration of treatment and the admission pain score.
da2$duration <- as.Date(da2$`Discharge Date`) - as.Date(da2$`Admission Date`)
da2$`Admission Pain` <- as.factor(da2$`Admission Pain`)
ggplot(subset(da2, `ROMS ID` != '2537'), aes(x=duration, color=`Admission Pain`)) +
    geom_density()+
    geom_line(stat="density")+
    scale_colour_brewer(palette = "Spectral")+
    facet_wrap(pain_effect~.)+
    labs( x=("treatment duration (days)"), y = ("density"), title= ("Distribution of treatment duration
    theme(plot.title = element_text(size =12))
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```

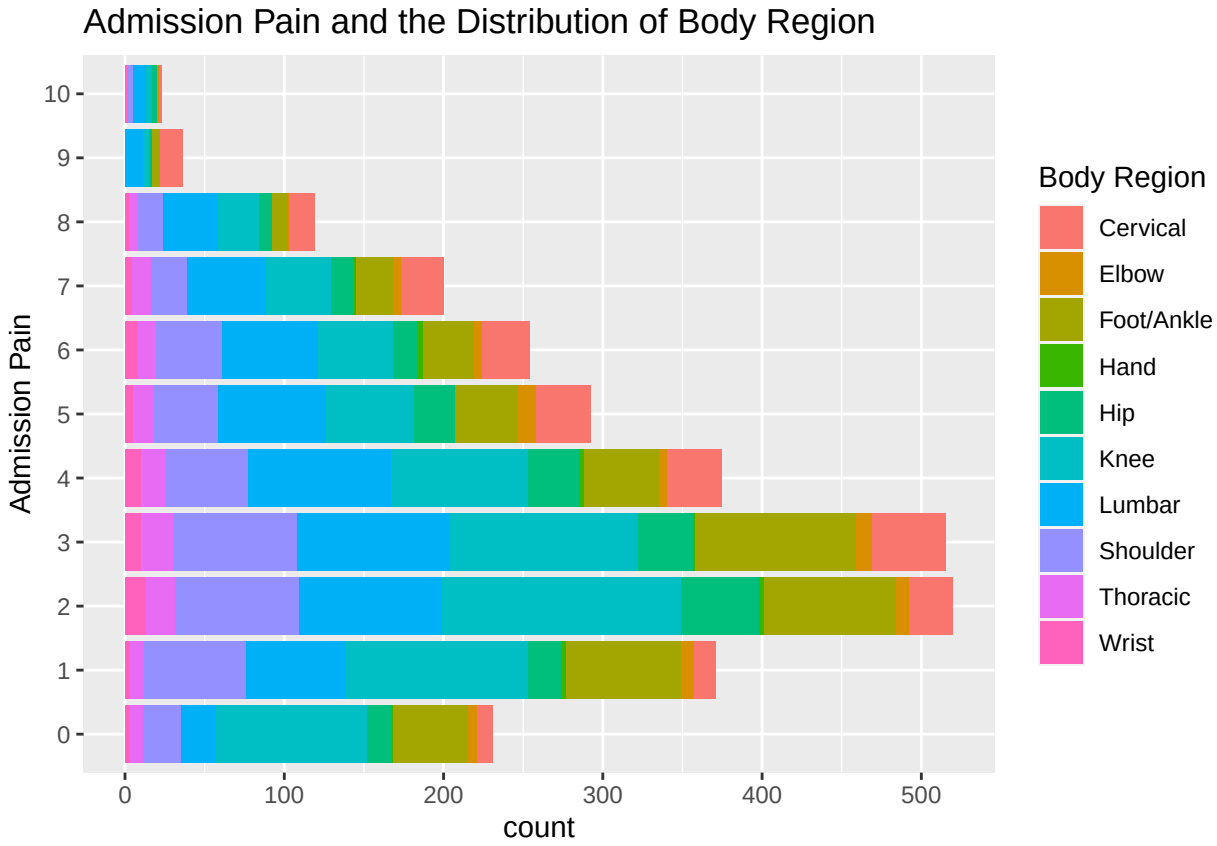



```
ggplot(subset(da2, `ROMS ID` != '2537'), aes(x=duration, color=`Admission Pain`)) +
    geom_density()+
    geom_line(stat="density")+
    scale_colour_brewer(palette = "Spectral")
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



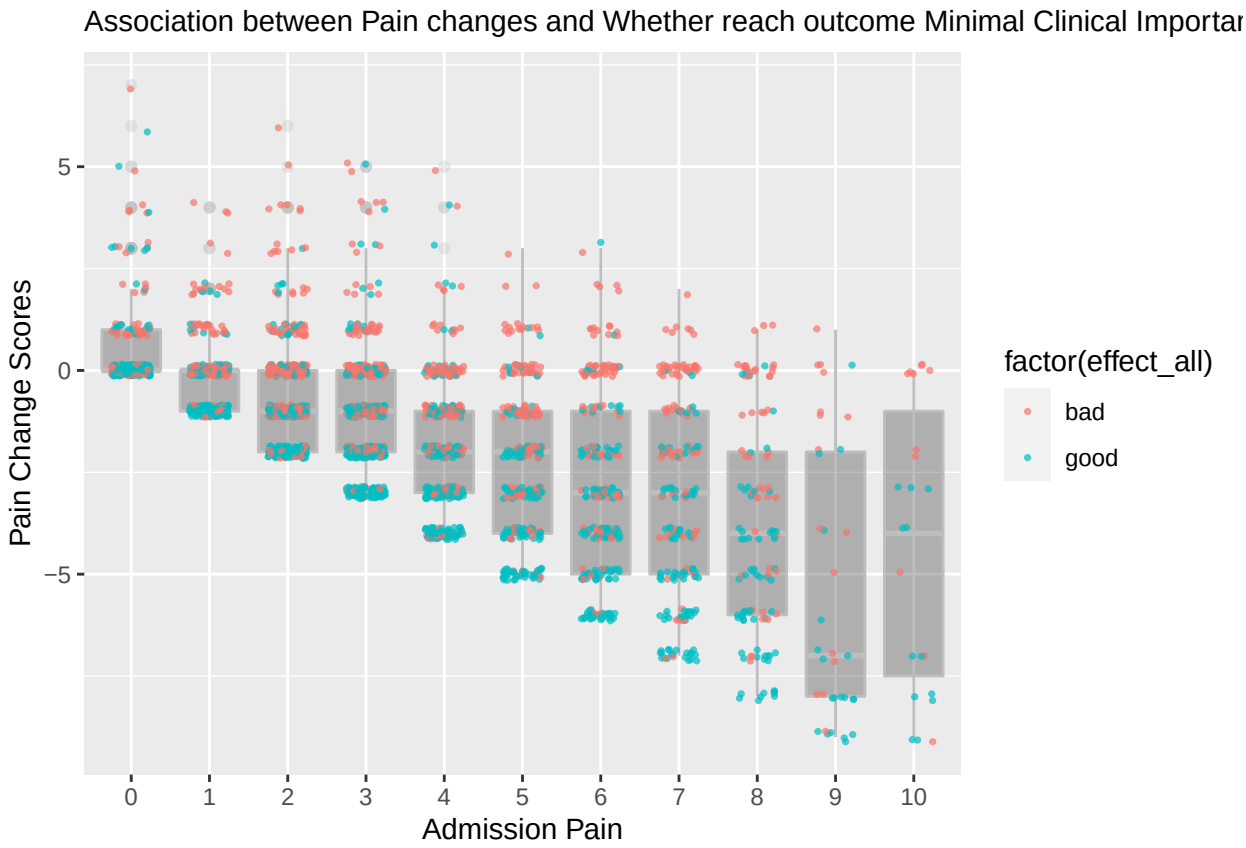

```
ggplot(da2, mapping = aes(x = `Admission Pain`, fill = `Body Region`)) +
    geom_bar(position="stack")+coord_flip() + labs(title = ("Admission Pain and the Distribution of Body
```




```
# pain changes scores vs admission pain scores. green color means the outcome improvement reaches the o
# pain relief and whether the outcome reaches minimal clinical important difference are positively asso

ggplot(da2, aes(x = `Admission Pain`, y = `Pain Change Scores`))+ geom_boxplot(col = "gray", alpha = 0.3
    theme(plot.title = element_text(size =11))
```

```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```


Association between Pain changes and Whether reach outcome Minimal Clinical Importar
Pain Change Scores
5
0
−5
0
1
2
3
4
5
6
7
8
9
10
Admission Pain
factor(effect_all)
bad
good