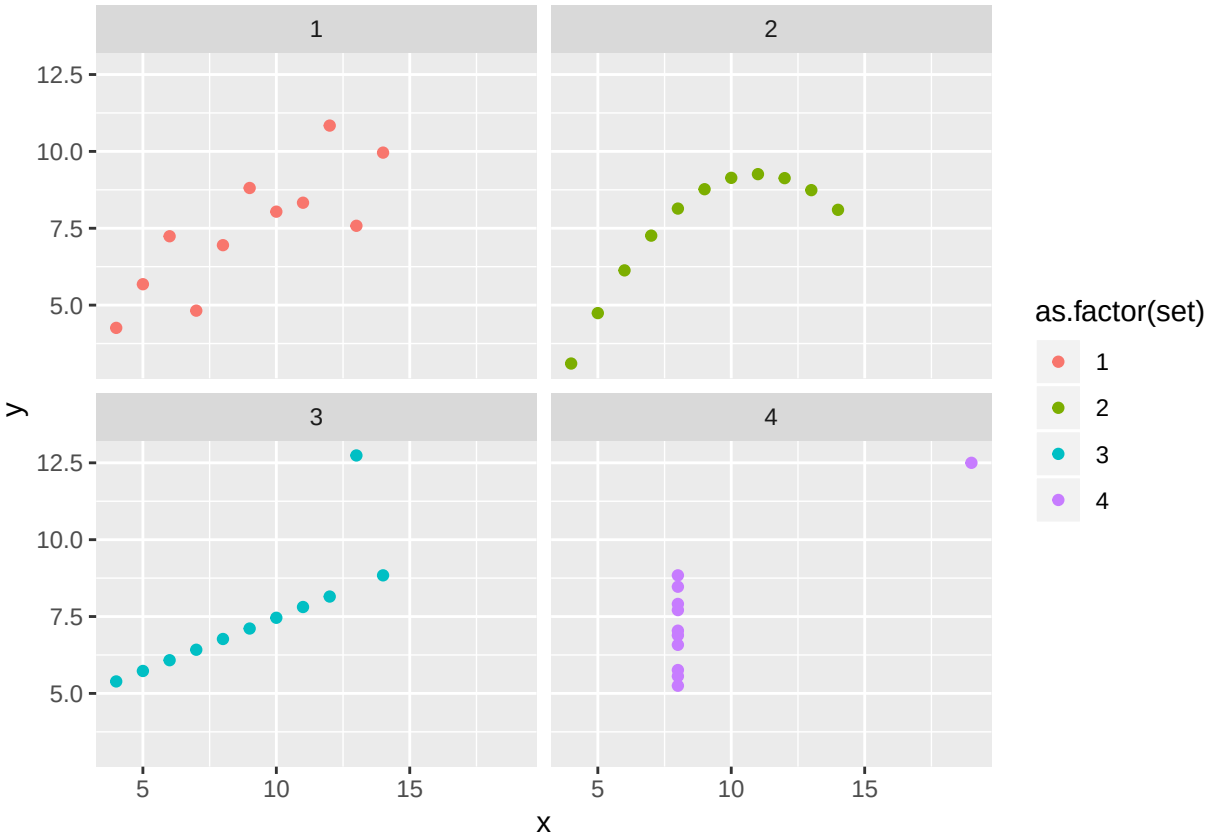
HW10

Oksana Ivanova

```
set.seed(42)
library(dplyr)
library(ggplot2)
data <- readRDS("~/Downloads/anscombe.rds")
head(data)
```

```
##    x    y set
## 1 10 8.04   1
## 2  8 6.95   1
## 3 13 7.58   1
## 4  9 8.81   1
## 5 11 8.33   1
## 6 14 9.96   1
```

```
data %>% ggplot(aes(x, y, color = as.factor(set))) +
          geom_point() +
          facet_wrap(.~set)
```




```
df1 <- data %>%
  group_by(set) %>%
  mutate(mean.x = mean(x), sd.x = sd(x))
df2 <- data %>%
  group_by(set) %>%
```

```
  mutate(mean.y = mean(y), sd.y = sd(y))
```

```
data <- merge(df1, df2)
head(data)
```

```
##    x    y set mean.x     sd.x   mean.y     sd.y
## 1 10 7.46   3      9 3.316625 7.500000 2.030424
## 2 10 8.04   1      9 3.316625 7.500909 2.031568
## 3 10 9.14   2      9 3.316625 7.500909 2.031657
## 4 11 7.81   3      9 3.316625 7.500000 2.030424
## 5 11 8.33   1      9 3.316625 7.500909 2.031568
## 6 11 9.26   2      9 3.316625 7.500909 2.031657
```

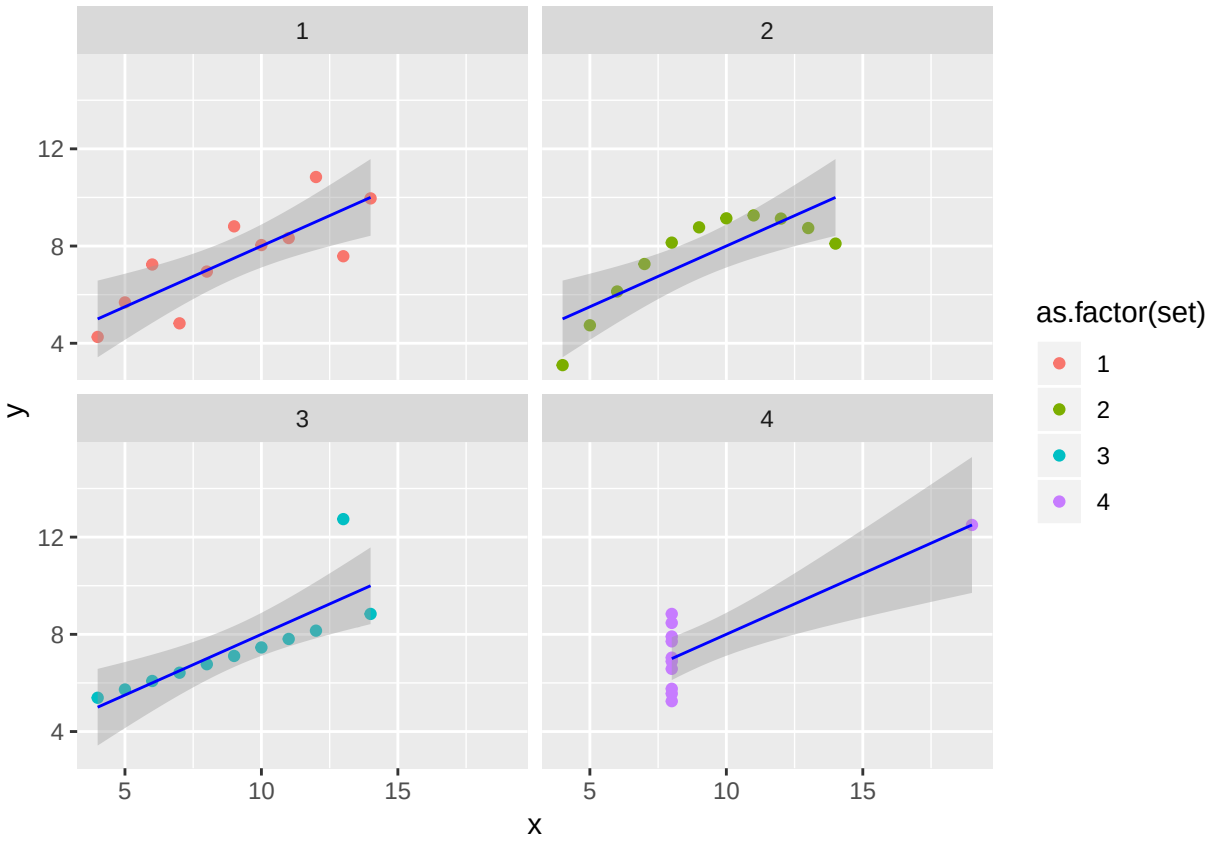
```
library(plyr)
sdata <- ddply(data, c("set"), summarise,
               mean.x = mean(x),
               mean.y = mean(y),
               sd.x = sd(x),
               sd.y = sd(y)
)
sdata
```

```
##   set mean.x   mean.y     sd.x     sd.y
## 1   1      9 7.500909 3.316625 2.031568
## 2   2      9 7.500909 3.316625 2.031657
## 3   3      9 7.500000 3.316625 2.030424
## 4   4      9 7.500909 3.316625 2.030579
```

```
ddply(data, "set", summarise, corr=cor(x, y),
      corspm = cor(x,y, method = "spearman"),
      p.value = cor.test(x,y)$p.value)
```

```
##   set      corr    corspm     p.value
## 1   1 0.8164205 0.8181818 0.002169629
## 2   2 0.8162365 0.6909091 0.002178816
## 3   3 0.8162867 0.9909091 0.002176305
## 4   4 0.8165214 0.5000000 0.002164602
```

```
data %>% ggplot(aes(x, y, color = as.factor(set))) +
  geom_point() +
  geom_smooth(color="blue", size = 0.5, method="lm") +
  facet_wrap(.~set)
```


1
2
3
4
12
8
4
5
10
15
y
x
as.factor(set)
1
2
3
4