Cor&lm

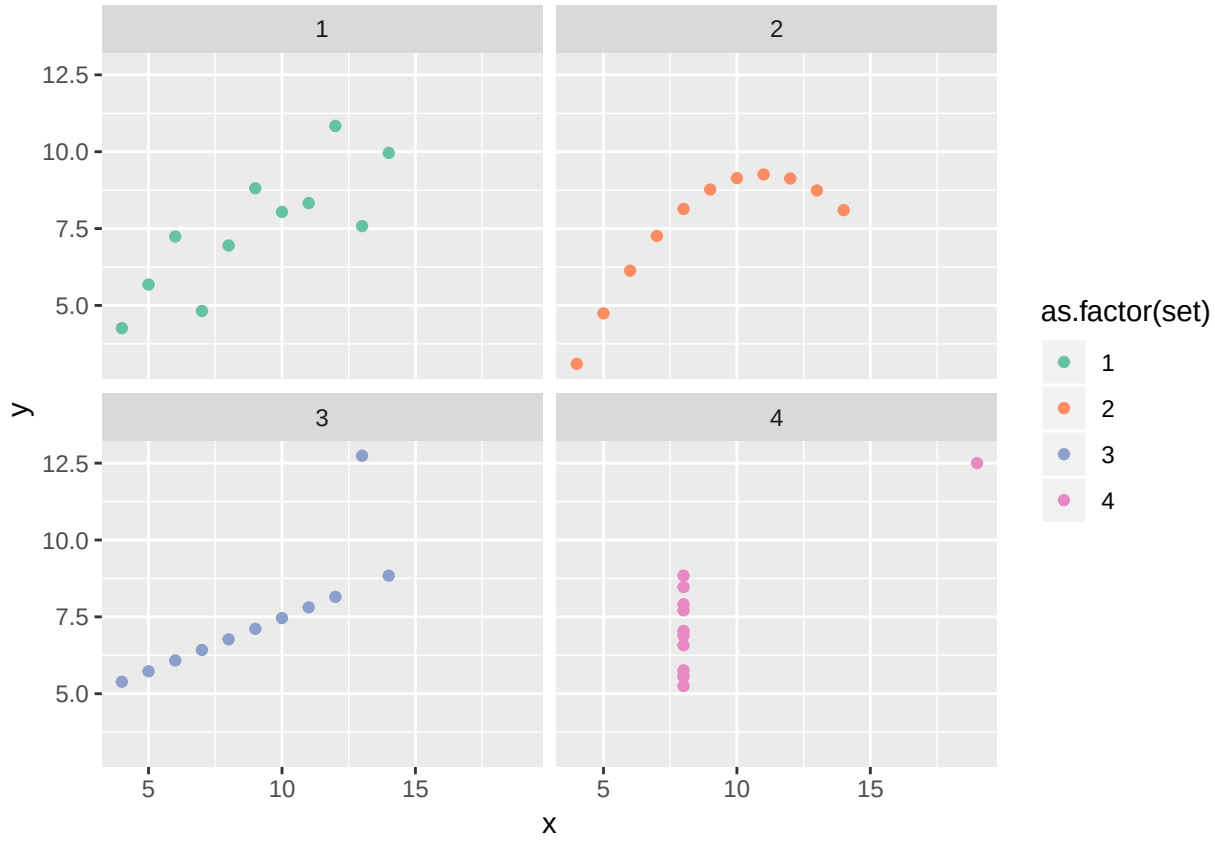
Natalia

29 May 2019

Task1

```
set.seed(42)

data <- readRDS('C://Users//Natalia//Desktop//ITMO//R//R_task#8//anscombe.rds')
View(data)
```

```
library(ggplot2)
ggplot(data = data, aes(x, y, color = as.factor(set))) +
  geom_point() +
  scale_color_brewer(type = "qual", palette = "Set2") +
  facet_wrap(.~set)
```




```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
```

```
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
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)
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
## 7  12 10.84   1      9 3.316625 7.500909 2.031568
## 8  12  8.15   3      9 3.316625 7.500000 2.030424
## 9  12  9.13   2      9 3.316625 7.500909 2.031657
## 10 13 12.74   3      9 3.316625 7.500000 2.030424
## 11 13  7.58   1      9 3.316625 7.500909 2.031568
## 12 13  8.74   2      9 3.316625 7.500909 2.031657
## 13 14  8.10   2      9 3.316625 7.500909 2.031657
## 14 14  8.84   3      9 3.316625 7.500000 2.030424
## 15 14  9.96   1      9 3.316625 7.500909 2.031568
## 16 19 12.50   4      9 3.316625 7.500909 2.030579
## 17  4  3.10   2      9 3.316625 7.500909 2.031657
## 18  4  4.26   1      9 3.316625 7.500909 2.031568
## 19  4  5.39   3      9 3.316625 7.500000 2.030424
## 20  5  4.74   2      9 3.316625 7.500909 2.031657
## 21  5  5.68   1      9 3.316625 7.500909 2.031568
## 22  5  5.73   3      9 3.316625 7.500000 2.030424
## 23  6  6.08   3      9 3.316625 7.500000 2.030424
## 24  6  6.13   2      9 3.316625 7.500909 2.031657
## 25  6  7.24   1      9 3.316625 7.500909 2.031568
## 26  7  4.82   1      9 3.316625 7.500909 2.031568
## 27  7  6.42   3      9 3.316625 7.500000 2.030424
## 28  7  7.26   2      9 3.316625 7.500909 2.031657
## 29  8  5.25   4      9 3.316625 7.500909 2.030579
## 30  8  5.56   4      9 3.316625 7.500909 2.030579
## 31  8  5.76   4      9 3.316625 7.500909 2.030579
## 32  8  6.58   4      9 3.316625 7.500909 2.030579
## 33  8  6.77   3      9 3.316625 7.500000 2.030424
## 34  8  6.89   4      9 3.316625 7.500909 2.030579
## 35  8  6.95   1      9 3.316625 7.500909 2.031568
## 36  8  7.04   4      9 3.316625 7.500909 2.030579
## 37  8  7.71   4      9 3.316625 7.500909 2.030579
## 38  8  7.91   4      9 3.316625 7.500909 2.030579
```

```
## 39  8  8.14   2      9 3.316625 7.500909 2.031657
## 40  8  8.47   4      9 3.316625 7.500909 2.030579
## 41  8  8.84   4      9 3.316625 7.500909 2.030579
## 42  9  7.11   3      9 3.316625 7.500000 2.030424
## 43  9  8.77   2      9 3.316625 7.500909 2.031657
## 44  9  8.81   1      9 3.316625 7.500909 2.031568
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

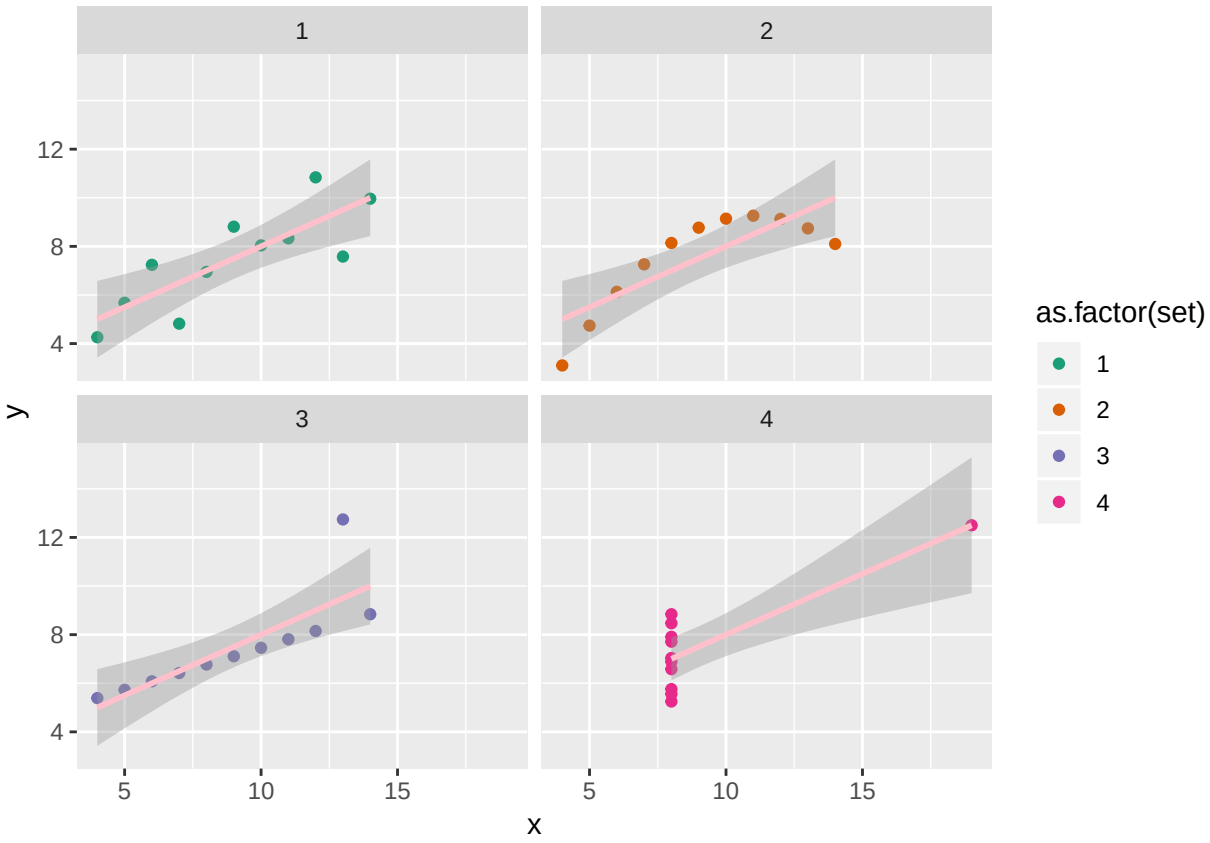
```
sdata <- ddply(data, c("set"), summarise,
               mean.x = mean(x),
               mean.y = mean(y),
               sd.x   = sd(x),
               sd.y = sd(y)

)
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
```

```
library(plyr)
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
```

```
ggplot(data = data, aes(x, y, color = as.factor(set))) +
  geom_point() +
  scale_color_brewer(type = "qual", palette = 2) +
  geom_smooth(color="pink", size = 1, method="lm") +
  facet_wrap(.~set)
```


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