

Coursera Statistical Inference Class

Pavel Gein

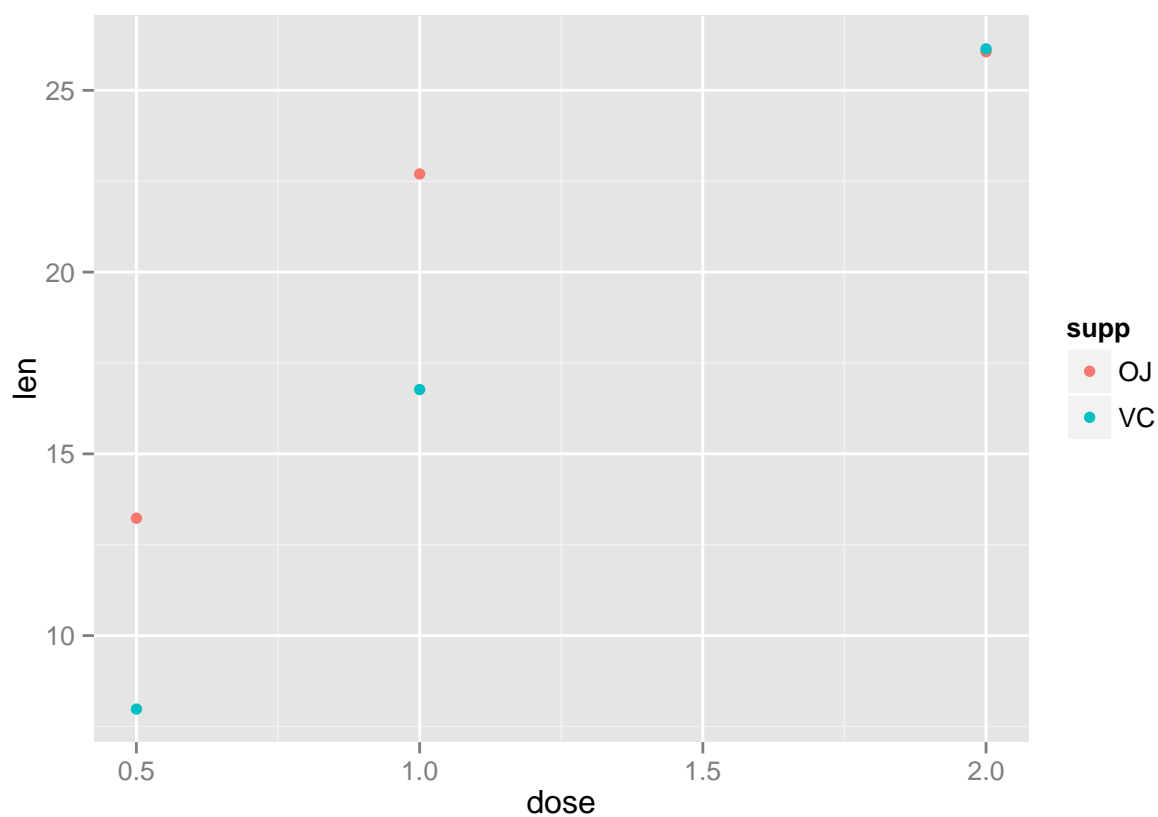
```
## Warning: package 'ggplot2' was built under R version 3.1.1
```

Load the data and look in summary of it

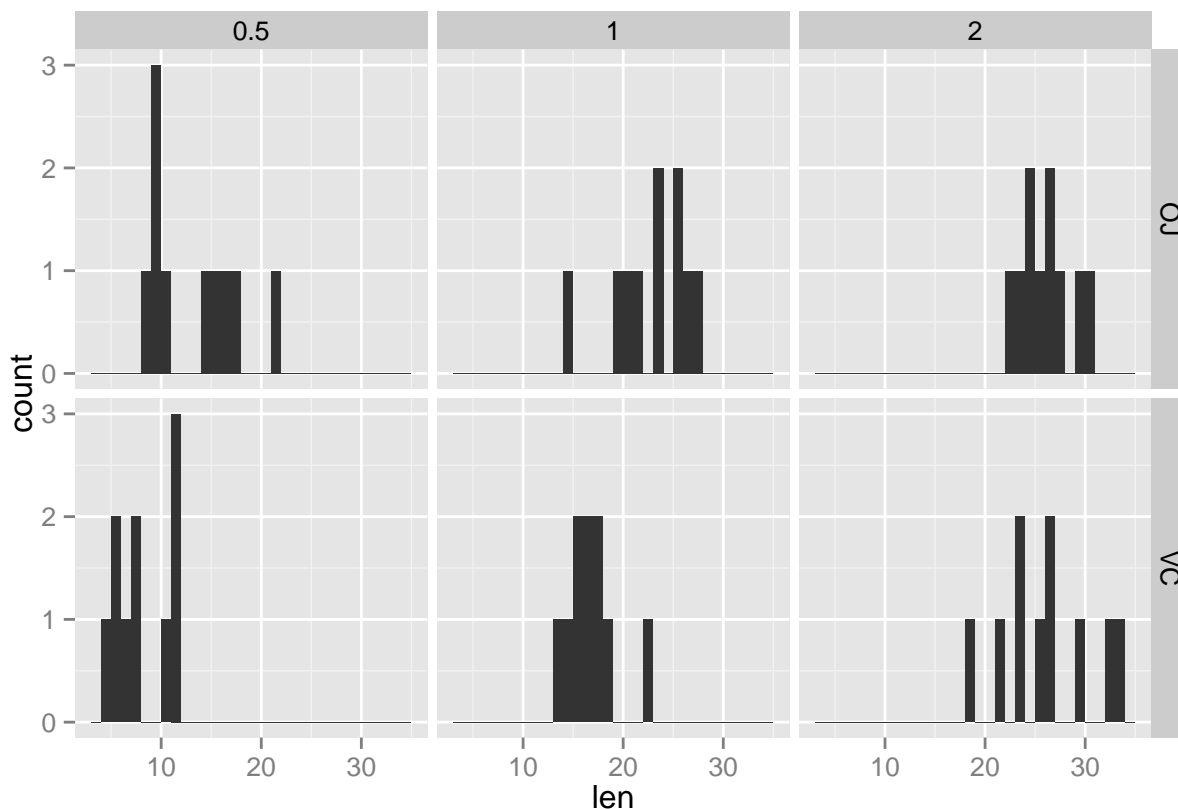
```
##      len      supp      dose
##  Min.   : 4.2    OJ:30    Min.   :0.50
##  1st Qu.:13.1    VC:30    1st Qu.:0.50
##  Median :19.2                Median :1.00
##  Mean   :18.8                Mean   :1.17
##  3rd Qu.:25.3                3rd Qu.:2.00
##  Max.   :33.9                Max.   :2.00
```

We can see that there is only 3 different doses and only 2 different supps.

Divide dataset into groups by dose and supp and calculate mean. Make a plot with different colors for each supp:



Investigate each group made by supp and dose (we've seen before a mean value of length).



Investigate is there a difference between mean value in each subgroup. We will use a T-test for this purpose. We assume that people doesn't change a dose and a supp.

```
## [1] "T-test for 0.5 dose:"
##
## Welch Two Sample t-test
##
## data: len by supp
## t = 3.17, df = 14.97, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719 8.781
## sample estimates:
## mean in group OJ mean in group VC
##      13.23      7.98
##
## [1] "T-test for 1 dose:"
##
## Welch Two Sample t-test
##
## data: len by supp
## t = 4.033, df = 15.36, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802 9.058
## sample estimates:
```

```
## mean in group OJ mean in group VC
##          22.70          16.77
##
## [1] "T-test for 2 dose:"
##
## Welch Two Sample t-test
##
## data: len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.798  3.638
## sample estimates:
## mean in group OJ mean in group VC
##          26.06          26.14
```

You can find a Rmd file [here](#)