

# DATA 605 Discussion Week 12

CUNY Spring 2021

Philip Tanofsky

24 April 2021

```
library(tidyverse)
library(ggplot2)
```

```
data("ToothGrowth")

tg <- ToothGrowth

dim(tg)
```

```
## [1] 60 3
```

```
summary(tg)
```

```
##      len      supp      dose
##  Min.   : 4.20   OJ:30   Min.    :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.    :2.000
```

From: <http://www.sthda.com/english/wiki/r-built-in-data-sets>

ToothGrowth data set contains the result from an experiment studying the effect of vitamin C on tooth growth in 60 Guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC)).

```
tg$dose.2 <- tg$dose^2

tg$supp <- as.character(tg$supp)
tg$supp[tg$supp == "OJ"] <- 0
tg$supp[tg$supp == "VC"] <- 1
tg$supp <- as.integer(tg$supp)

tg$supp.x.dose <- tg$supp * tg$dose

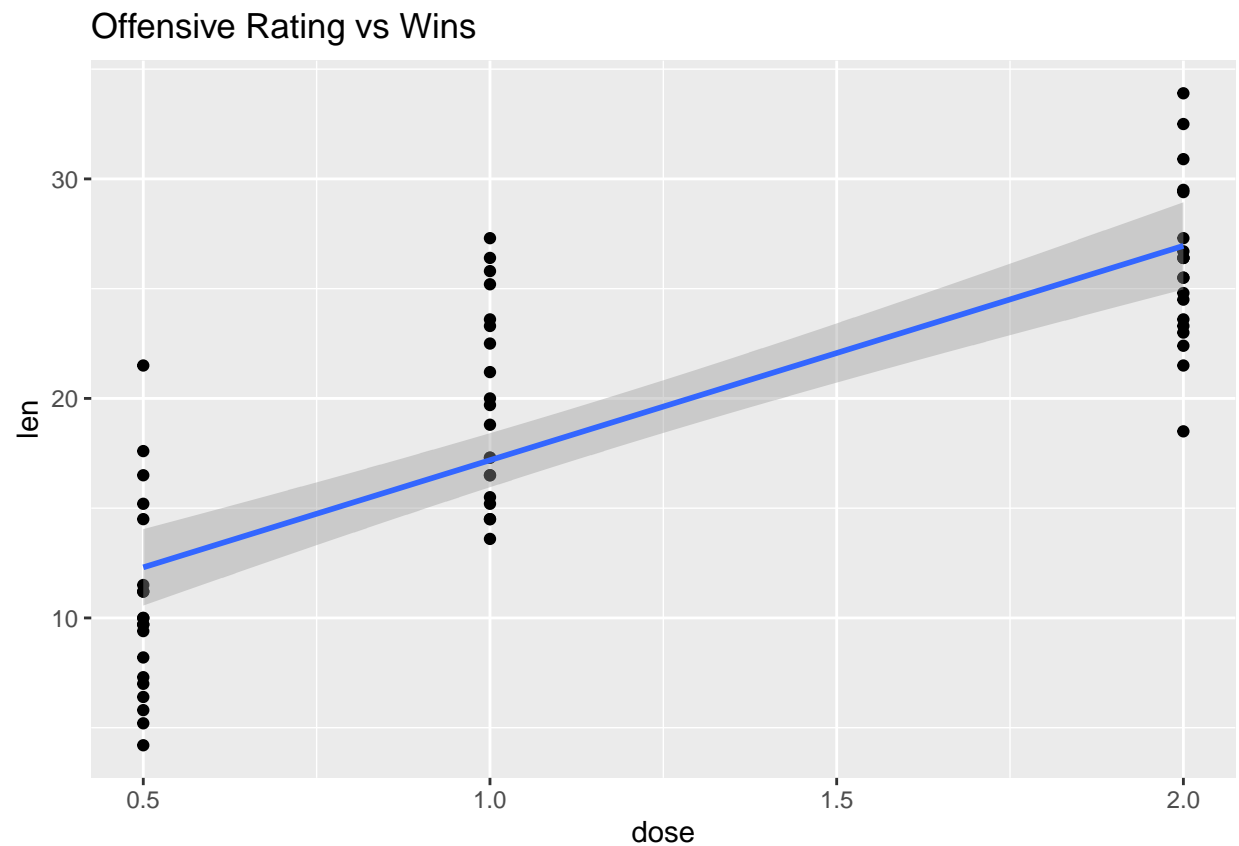
tg$supp <- as.factor(tg$supp)

summary(tg)
```

##	len	supp	dose	dose.2	supp.x.dose
##	Min. : 4.20	0:30	Min. :0.500	Min. :0.25	Min. :0.0000
##	1st Qu.:13.07	1:30	1st Qu.:0.500	1st Qu.:0.25	1st Qu.:0.0000
##	Median :19.25		Median :1.000	Median :1.00	Median :0.2500
##	Mean :18.81		Mean :1.167	Mean :1.75	Mean :0.5833
##	3rd Qu.:25.27		3rd Qu.:2.000	3rd Qu.:4.00	3rd Qu.:1.0000
##	Max. :33.90		Max. :2.000	Max. :4.00	Max. :2.0000

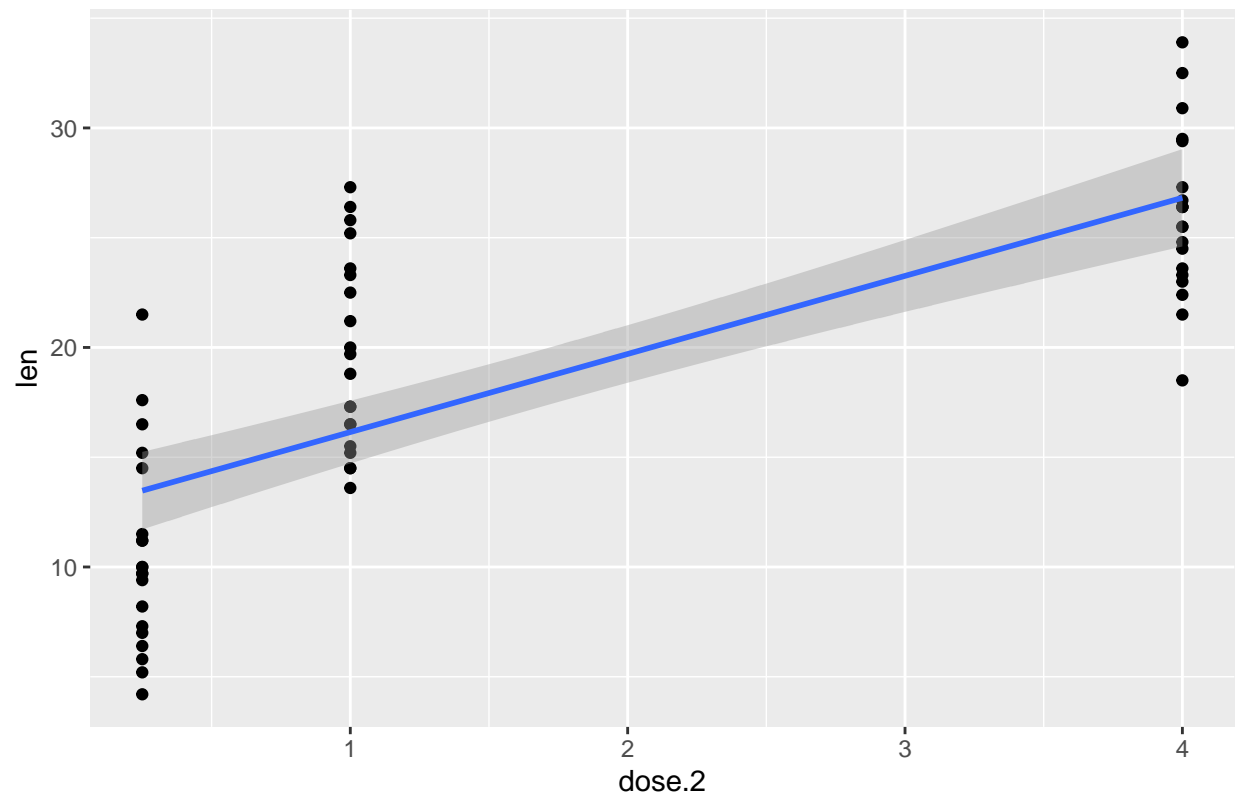
Plots

```
tg %>%
  ggplot(aes(x=dose, y=len)) +
  geom_point() +
  labs(title = 'Offensive Rating vs Wins') + geom_smooth(method='lm', formula= y~x)
```



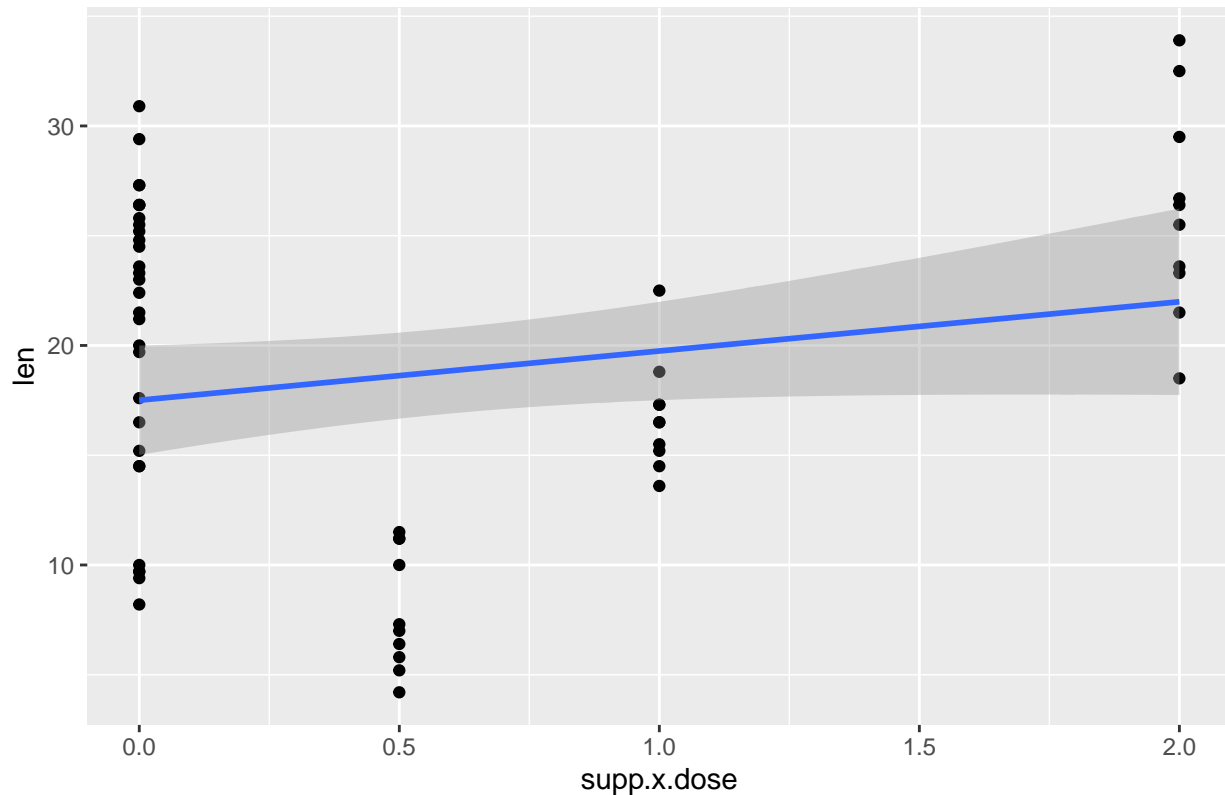
```
tg %>%
  ggplot(aes(x=dose.2, y=len)) +
  geom_point() +
  labs(title = 'Defensive Rating vs Wins') + geom_smooth(method='lm', formula= y~x)
```

Defensive Rating vs Wins



```
tg %>%
  ggplot(aes(x=supp.x.dose, y=len)) +
  geom_point() +
  labs(title = 'Rating Difference vs Wins') + geom_smooth(method='lm', formula= y~x)
```

## Rating Difference vs Wins



```
tg_lm <- lm(len ~ ., data=tg)
summary(tg_lm)
```

```
##
## Call:
## lm(formula = len ~ ., data = tg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.4104 -2.2975 -0.3182  2.8269  8.0232
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.637      2.909   0.563 0.575741
## supp1         -8.255      1.998  -4.133 0.000123 ***
## dose          28.203      5.339   5.283 2.24e-06 ***
## dose.2        -7.930      2.034  -3.898 0.000265 ***
## supp.x.dose     3.904      1.510   2.586 0.012398 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.647 on 55 degrees of freedom
## Multiple R-squared:  0.7881, Adjusted R-squared:  0.7727
## F-statistic: 51.14 on 4 and 55 DF,  p-value: < 2.2e-16
```