Tooth Growth Dataset Exploration

Maria Oros

2025-01-22

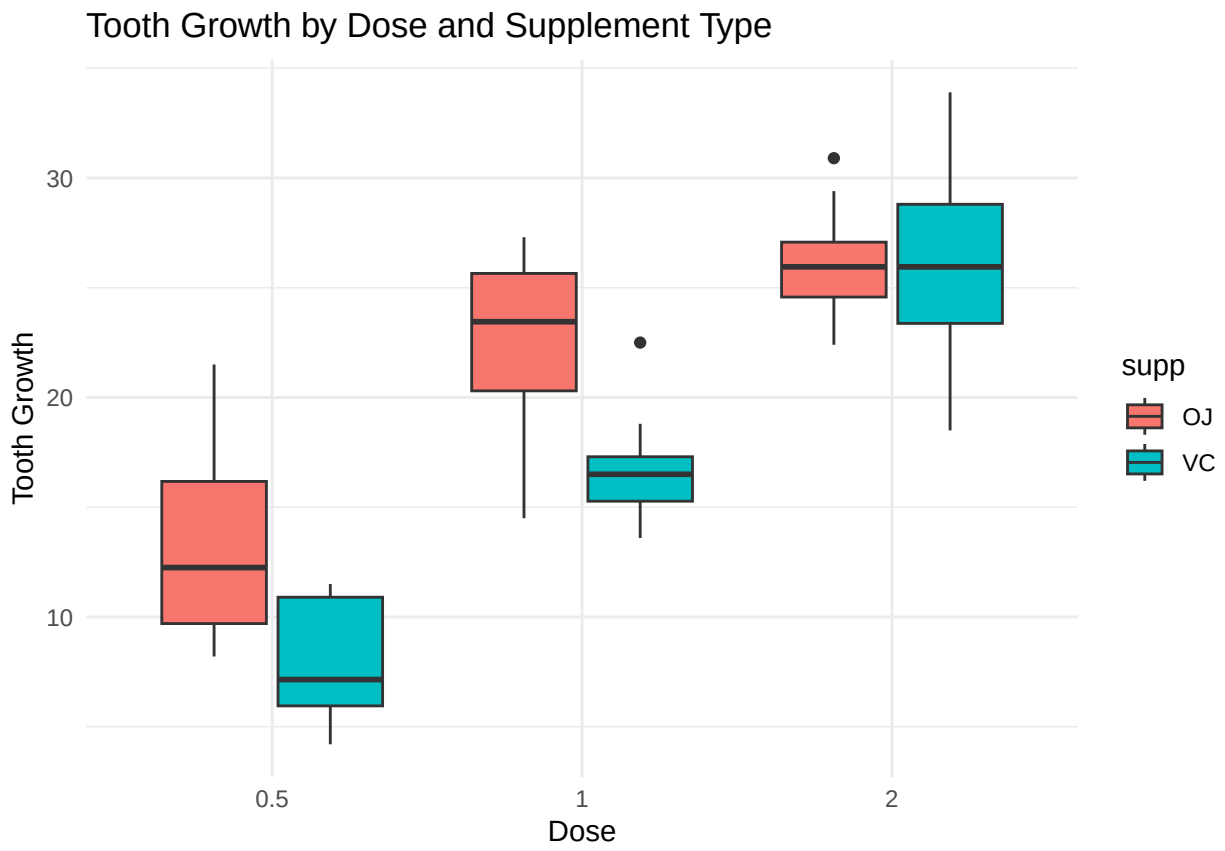
Data Loading

```
data("ToothGrowth")
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

Visualization

```
ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=supp)) +
  geom_boxplot() +
  labs(x = "Dose", y = "Tooth Growth", title = "Tooth Growth by Dose and Supplement Type") +
  theme_minimal()
```



Statistical Analysis

```
t_test_results <- lapply(unique(ToothGrowth$dose), function(dose_level) {
  data_subset <- subset(ToothGrowth, dose == dose_level)
  t.test(len ~ supp, data = data_subset)
})
print(t_test_results)
```

```
## [[1]]
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
##
##
## [[2]]
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
##
##
## [[3]]
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Normality Check

```
shapiro_test_results <- lapply(unique(ToothGrowth$dose), function(dose_level) {
  data_subset <- subset(ToothGrowth, dose == dose_level)
  shapiro.test(data_subset$len)
})
print(shapiro_test_results)
```

```
## [[1]]
##
##  Shapiro-Wilk normality test
##
## data:  data_subset$len
## W = 0.94065, p-value = 0.2466
##
##
## [[2]]
##
##  Shapiro-Wilk normality test
##
## data:  data_subset$len
## W = 0.93134, p-value = 0.1639
##
##
## [[3]]
##
```

```
##  Shapiro-Wilk normality test
##
## data:  data_subset$len
## W = 0.97775, p-value = 0.9019
```