Rainfall Example

David Gerard

2018-12-07

Objectives

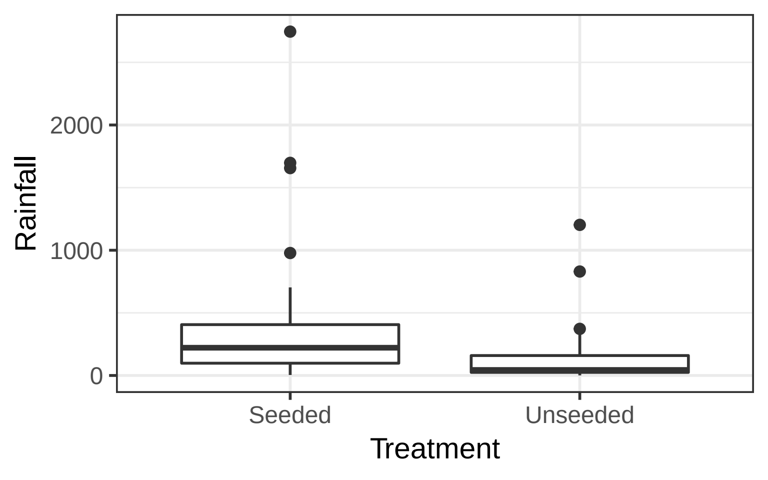
Here, we work through the rainfall analysis

Load in Data

```r
library(Sleuth3)
library(ggplot2)
data("case0301")
```

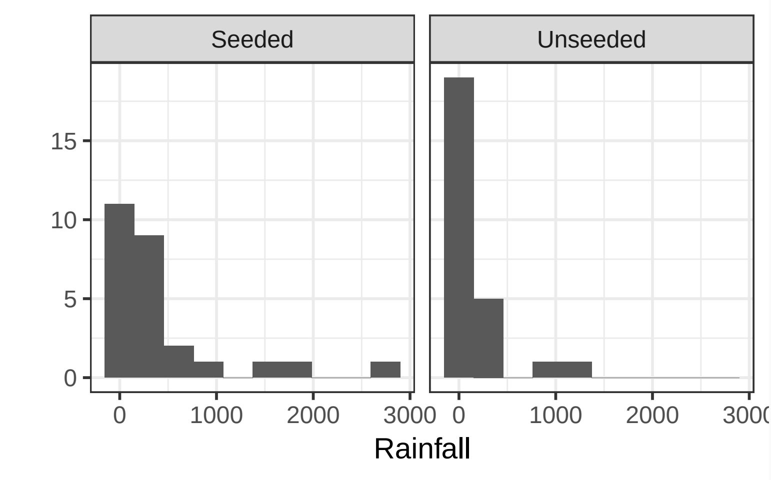
Rainfall EDA

```
qplot(x = Treatment, y = Rainfall,
      data = case0301, geom = "boxplot")
```



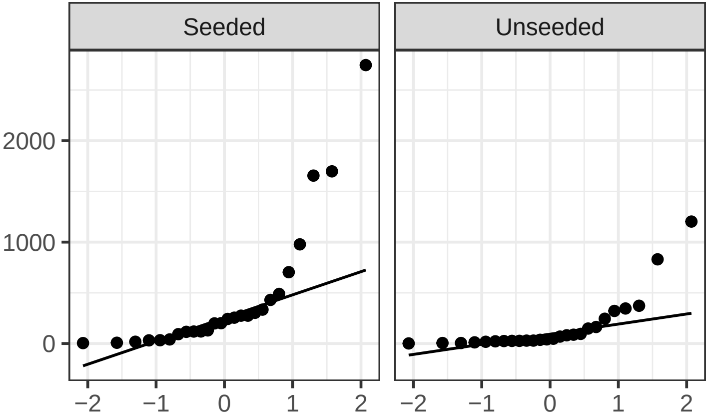
Rainfall EDA

```
qplot(x = Rainfall, facets = . ~ Treatment,
      data = case0301, geom = "histogram", bins = 10)
```



Rainfall EDA

```
qplot(sample = Rainfall, facets = . ~ Treatment,
      data = case0301, geom = "qq") +
  geom_qq_line()
```

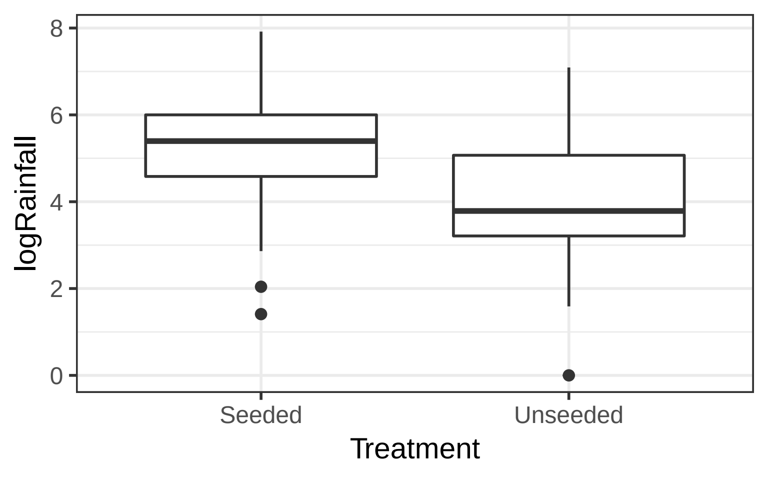


Apply Transformation

```
case0301$logRainfall <- log(case0301$Rainfall)
```

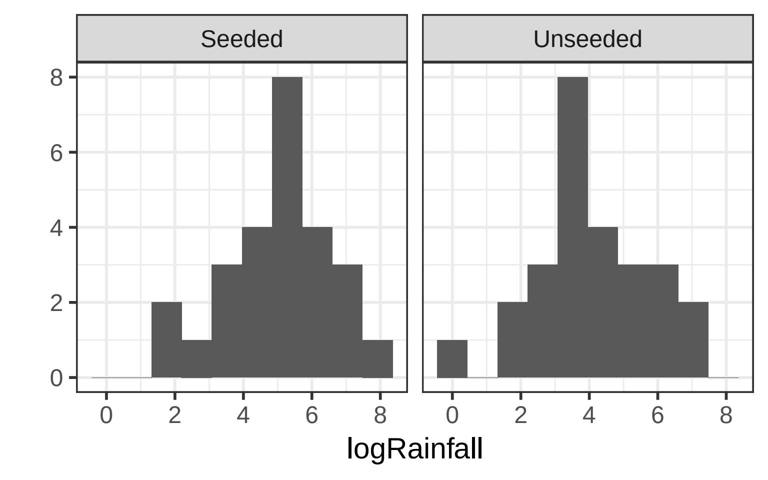
Rain EDA

```
qplot(x = Treatment, y = logRainfall,
      data = case0301, geom = "boxplot")
```



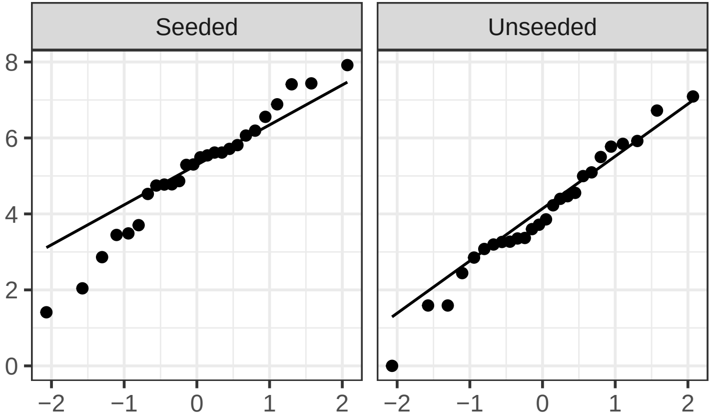
Rainfall EDA

```r
qplot(x = logRainfall, facets = . ~ Treatment,
      data = case0301, geom = "histogram", bins = 10)
```



Rainfall EDA

```
qplot(sample = logRainfall, facets = . ~ Treatment,
      data = case0301, geom = "qq") +
  geom_qq_line()
```



Interpretation

Posit a Model

- Z_i = rainfall on unseeded days.
- Y_i = **log** rainfall on unseeded days.
- Z_i^* = rainfall on seeded days.
- Y_i^* = **log** rainfall on seeded days.
- $Y_i^* = Y_i + \delta$.
- $Z_i^* = e^\delta Z_i$.
- e^δ is the multiplicative effect of seeding on rainfall.
- $e^\delta = 2$ means rainfall is twice as large on seeded days.
- $e^\delta = 3$ means rainfall is three times as large on seeded days.

Posit Hypotheses

- $H_0 : \delta = 0$
- $H_a : \delta \neq 0$

Run t-test

```
tout <- t.test(logRainfall ~ Treatment, data = case0301)
tout
```

```
##
##  Welch Two Sample t-test
##
## data:  logRainfall by Treatment
## t = 2.5, df = 50, p-value = 0.01
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  0.2408 2.0467
## sample estimates:
##   mean in group Seeded mean in group Unseeded
##                  5.134                  3.990
```

Estimate and Confidence Intervals on Original Scale

```r
exp(tout$estimate[1] - tout$estimate[2])
```

```
## mean in group Seeded
##                3.139
```

```r
exp(tout$conf.int)
```

```
## [1] 1.272 7.742
## attr(,"conf.level")
## [1] 0.95
```

Conclusion

- We estimate that that seeding results in a 3.1 factor increase in rainfall (p-value 0.01, 95% confidence interval of 1.3 to 7.7).
- Note the causal language because this is a randomized experiment. I will deduct many points if you use causal language in an observational study.