

# ANOVA, single and multiple factor experiments

EXPERIMENTAL DESIGN IN R



**Kaelen Medeiros**  
Data Scientist

# ANOVA

- Used to compare 3+ groups
- An omnibus test:
  - won't know which groups' means are different without additional post hoc testing
- Two ways to implement in R:

*#one*

```
model_1 <- lm(y ~ x, data = dataset)
anova(model_1)
```

*#two*

```
aov(y ~ x, data = dataset)
```

# Single factor experiments

```
model_1 <- lm(y ~ x)
```

- **y** = outcome variable
  - Tensile strength of different cotton fabrics
- **x** = explanatory factor variable
  - Percent cotton in the fabric

# Multiple factor experiments

```
model2 <- lm(y ~ x + r + s + t)
```

- `y` = outcome
  - ToothGrowth length
- `x` , `r` , `s` , `t` = possible explanatory factor variables
  - How much vitamin C & delivery method

# Intro to Lending Club data

- Lending Club is a U.S. based peer-to-peer loan company.
  - Data is openly available on Kaggle
- Includes all loans issued from 2007-2015
- Big!
  - 890k observations and 75 variables

**Let's practice!**  
EXPERIMENTAL DESIGN IN R

# Model validation

EXPERIMENTAL DESIGN IN R



**Kaelen Medeiros**

Data Scientist

# Pre-modeling EDA

- Mean and variance of outcome by variable of interest

```
lendingclub %>% summarise(median(loan_amnt),  
                           mean(int_rate),  
                           mean(annual_inc))  
  
lendingclub %>% group_by(verification_status) %>%  
  summarise(mean(funded_amnt),  
            var(funded_amnt))
```

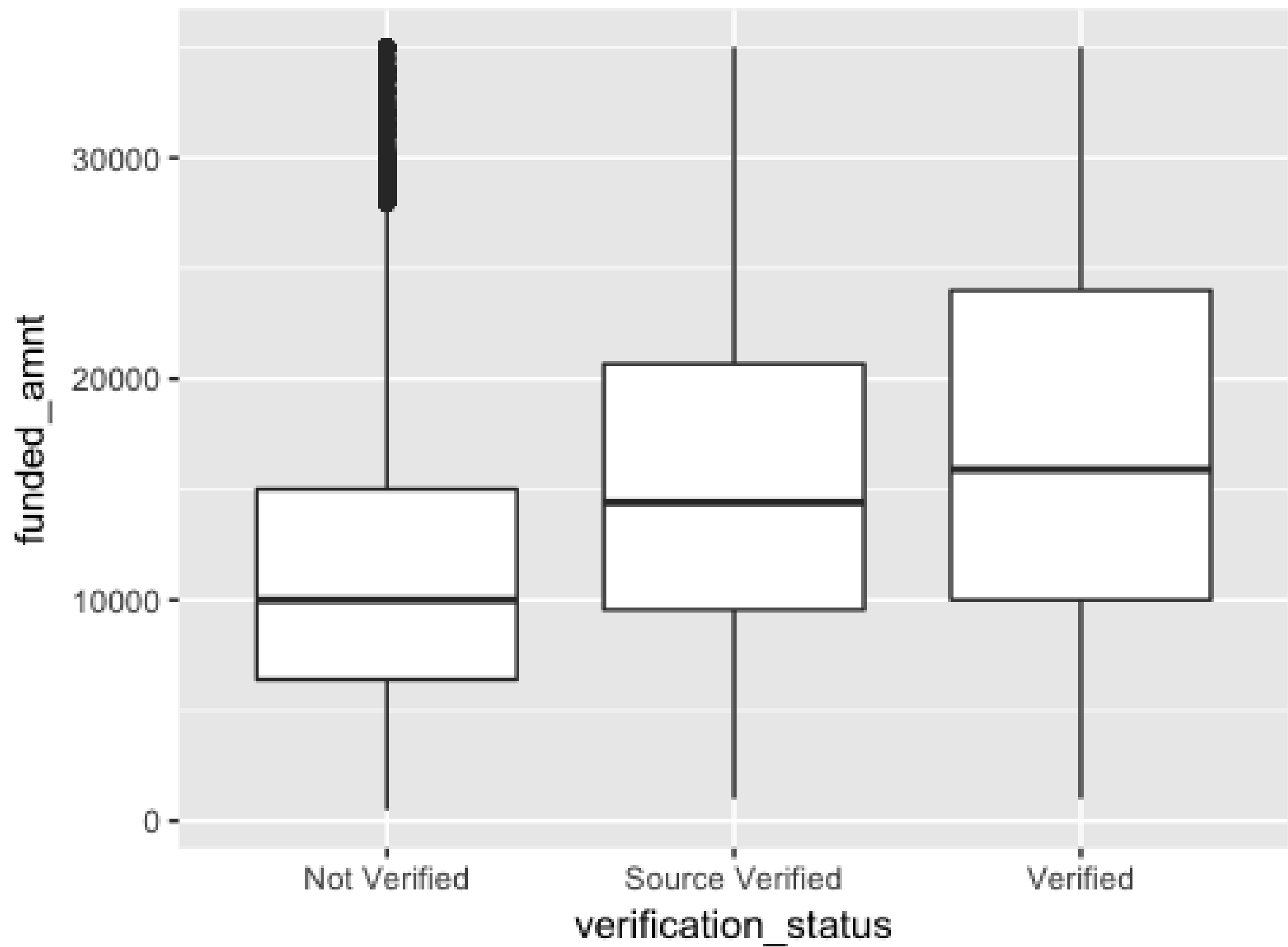
```
# A tibble: 3 x 3  
verification_status `mean(funded_amnt)` `var(funded_amnt)`  
<chr>              <dbl>          <dbl>  
1 Not Verified      114.15        349.41953  
2 Source Verified   156.14        723.53265  
3 Verified          166.08        848.54561
```



# Pre-modeling EDA continued

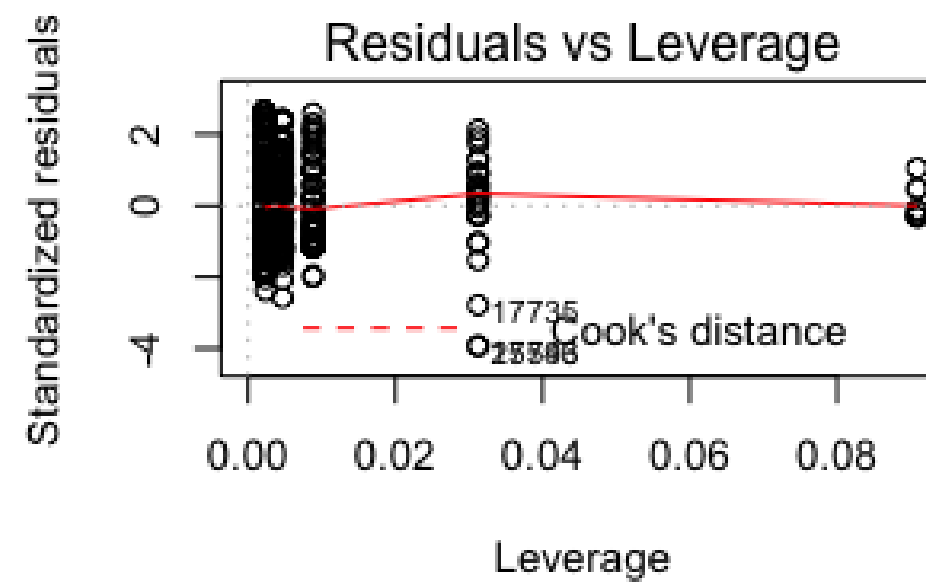
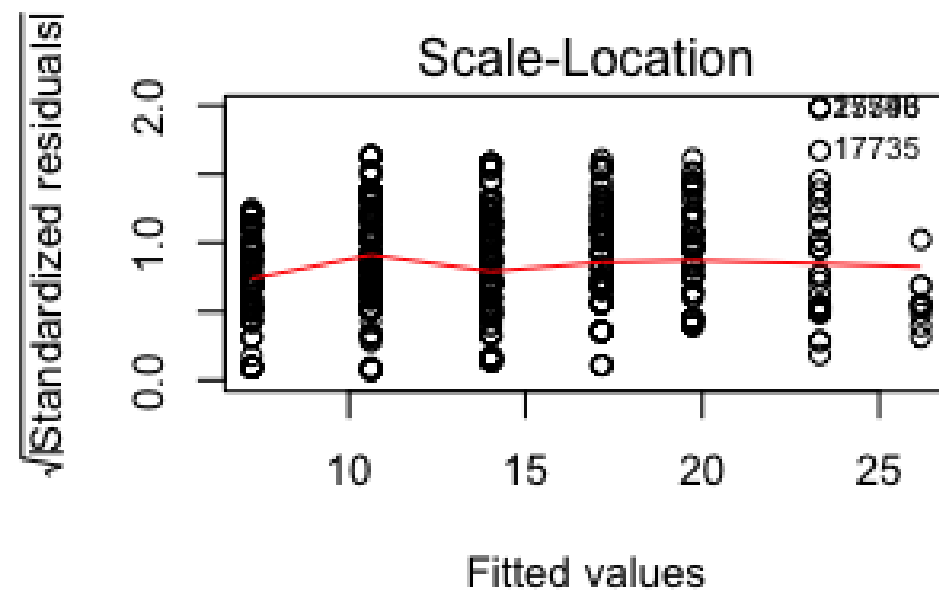
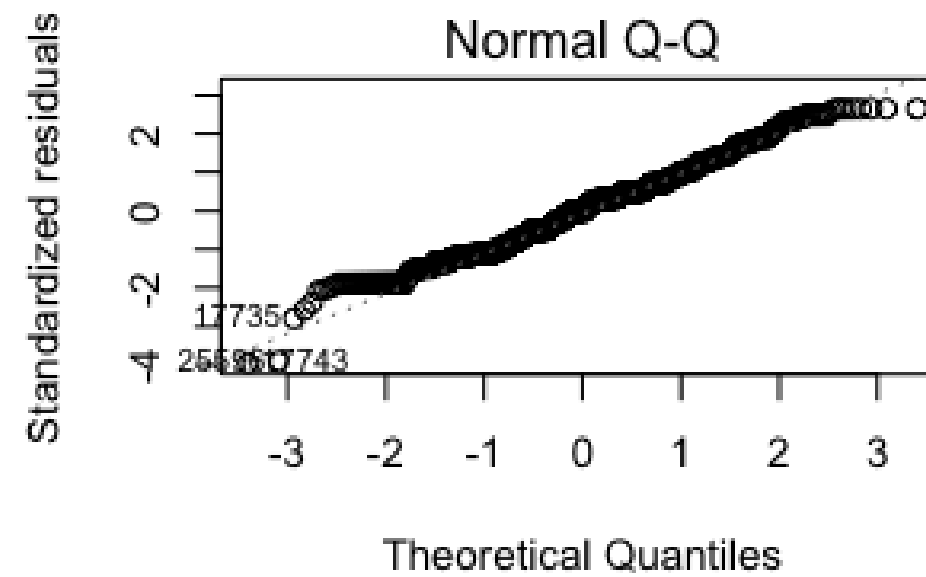
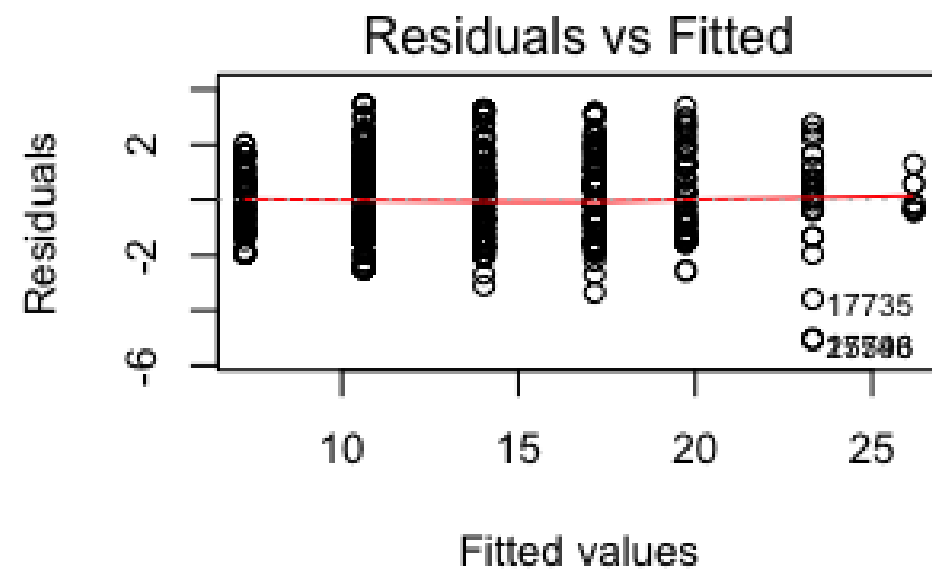
- Boxplot of outcome (y-axis) by variable of interest (x-axis).

```
ggplot(data = lendingclub,  
       aes(x = verification_status, y = funded_amnt)) +  
  geom_boxplot()
```



# Post-modeling model validation

- Residual plot
- QQ-plot for normality
- Test ANOVA assumptions
  - Homogeneity of variances
- Try non-parametric alternatives to ANOVA



**Let's practice!**  
EXPERIMENTAL DESIGN IN R

# A/B testing

## EXPERIMENTAL DESIGN IN R



**Kaelen Medeiros**  
Data Scientist

# A/B testing

- A type of controlled experiment with only two variants of something, for example:
  - 1 word different in a marketing email
  - Red 'buy' button on a website vs. blue button
  - How many consumers click through to create an account based on two different website headers?

# Power and sample size in A/B tests

- Calculate sample size, given some power, significance level, and effect size
- Run your A/B test until you attain the sample size you calculated



# Lending Club A/B test

LENDING CLUB  
Borrow money!!!!!!

Lorem ipsum dolor sit amet, cu sensibus  
referrentur per. Qui illum neglegentur et,  
cetero volutpat vix in, has nulla nostra id.  
Et mea feugait docendi appellantur, at  
est verear veritus prodesset, mea ea  
choro fastidii mandamus. Quod  
dissentiunt vel et, mel vitae aperiam id,  
has ad dico scripta euripidis. Mea ex nisl  
voluptaria, qui porro viderer hendrerit ex.  
Per et munere causae, at dolorem  
admodum his.

Solum munere assueverit duo et, percipit  
forensibus sed ut. Nec ex liber graece.  
lisque accommodare no sit, mel ea inani  
liberavisse. Omittam delicatissimi pro ea.  
Aperiri noluisse deseruisse ut mel, at  
euismod debitis eam. Id illum ullum  
periculis eam.

APPLY

LENDING CLUB  
Borrow money!!!!!!

Lorem ipsum dolor sit amet, cu sensibus  
referrentur per. Qui illum neglegentur et,  
cetero volutpat vix in, has nulla nostra id.  
Et mea feugait docendi appellantur, at  
est verear veritus prodesset, mea ea  
choro fastidii mandamus. Quod  
dissentiunt vel et, mel vitae aperiam id,  
has ad dico scripta euripidis. Mea ex nisl  
voluptaria, qui porro viderer hendrerit ex.  
Per et munere causae, at dolorem  
admodum his.

Solum munere assueverit duo et, percipit  
forensibus sed ut. Nec ex liber graece.  
lisque accommodare no sit, mel ea inani  
liberavisse. Omittam delicatissimi pro ea.  
Aperiri noluisse deseruisse ut mel, at  
euismod debitis eam. Id illum ullum  
periculis eam.

APPLY

**Let's practice!**  
EXPERIMENTAL DESIGN IN R