A/B TESTING IN R

Power Analyses

Page Piccinini
Instructor

What are power analyses? - Cambridge Dictionary of Statistics

Power

Significance level

Effect size

What are power analyses? - Cambridge Dictionary of Statistics

Power:

- The probability of **rejecting the null hypothesis** when it is false.
- It is also the basis of procedures for **estimating the sample size** needed to detect an **effect of a particular magnitude**
- Power gives a method of discriminating between competing tests of the same hypothesis, the test with the higher power being preferred.

What are power analyses? - Cambridge Dictionary of Statistics

Significance level:

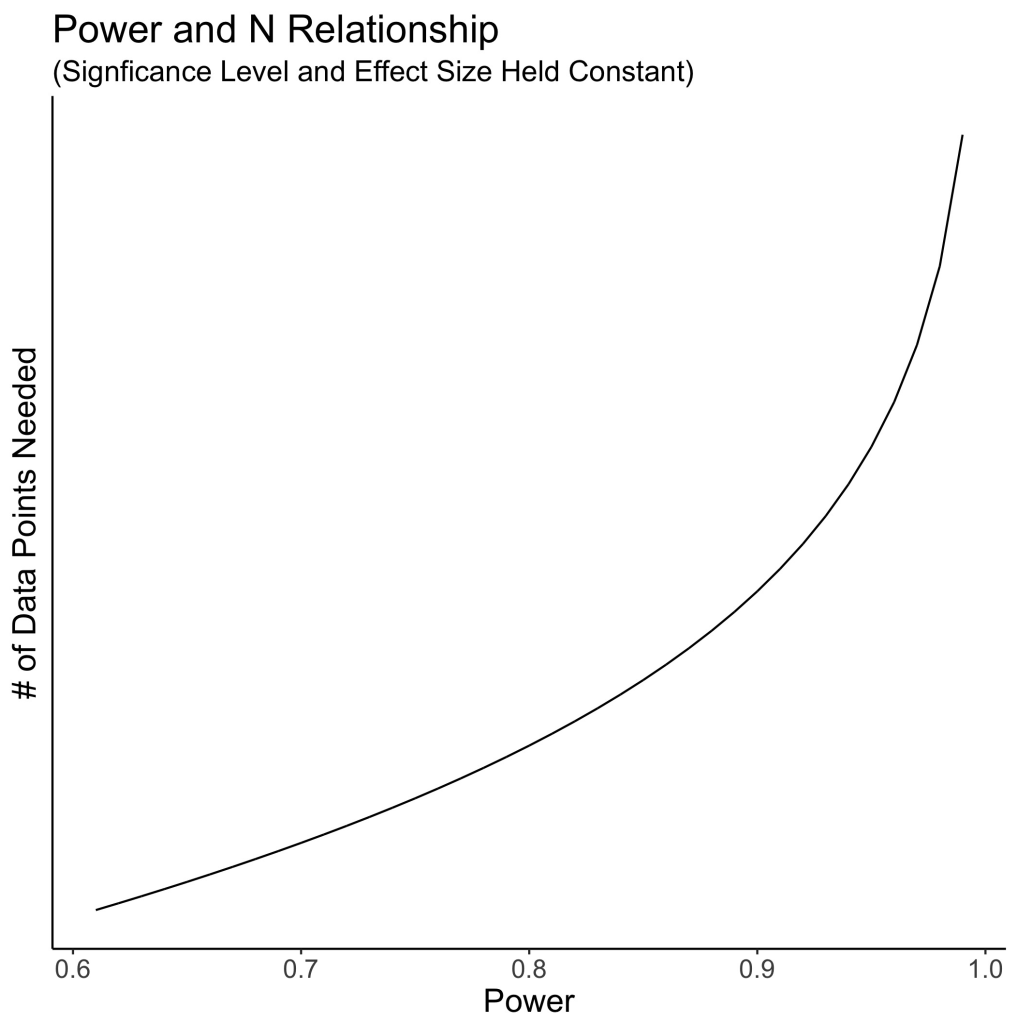
- The **level of probability** at which it is agreed that the **null hypothesis will be rejected**.
- Conventionally set at 0.05.

What are power analyses? - Cambridge Dictionary of Statistics

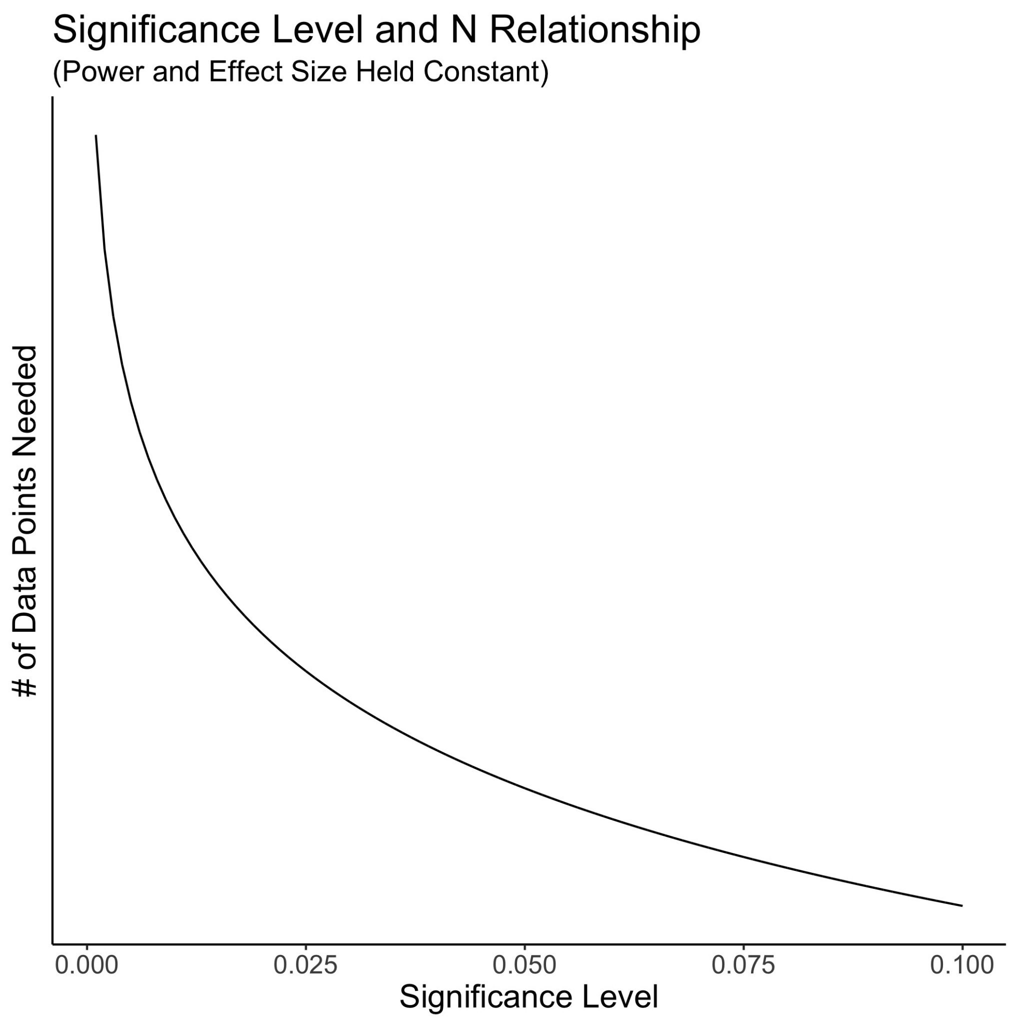
Effect size:

- Most commonly the **difference between the control group and experimental group** population means of a response variable **divided by the assumed common population standard deviation**.
- Estimated by the difference of the sample means in the two groups divided by a pooled estimate of the assumed common standard deviation.

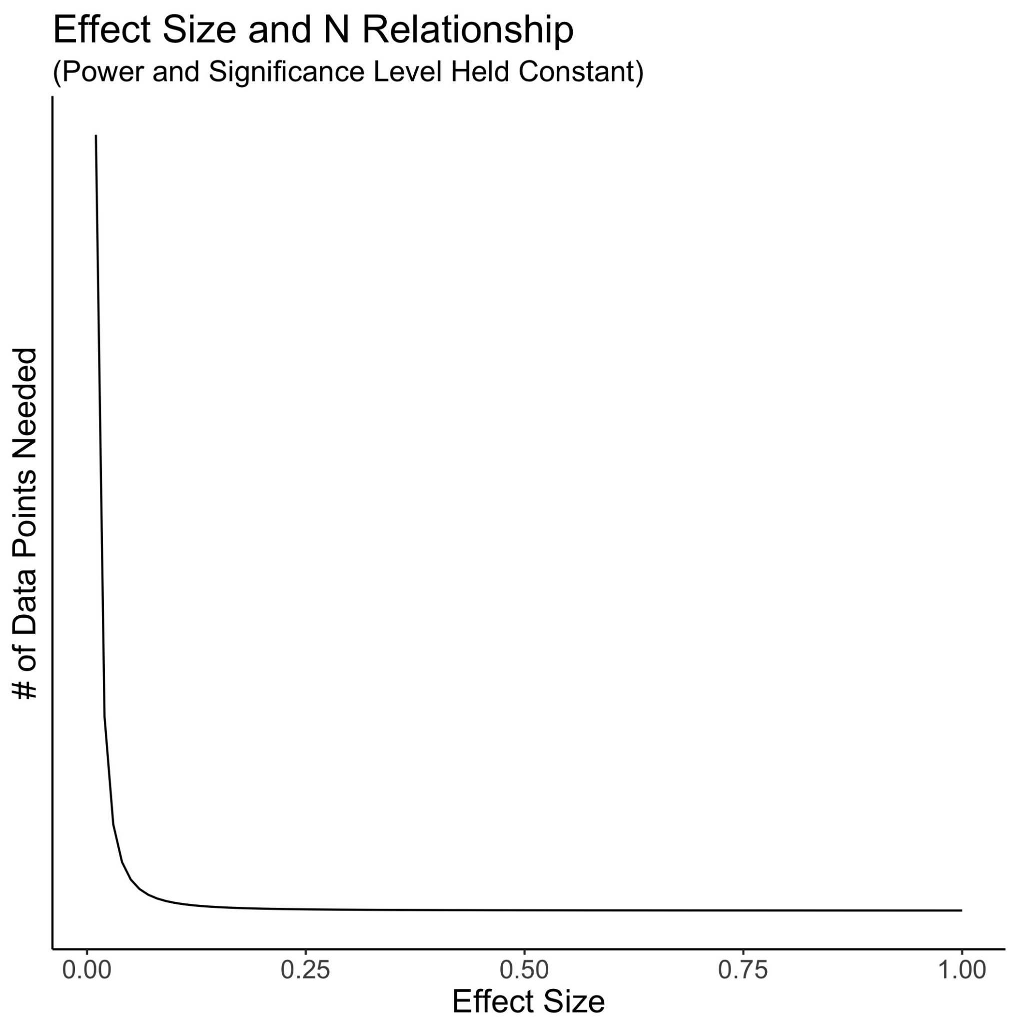
Power analysis relationships



Power analysis relationships



Power analysis relationships



Power analysis in R: T-Test

```
library(pwr)

pwr.t.test(

                        )
```

Power analysis in R: T-Test

```
library(pwr)

pwr.t.test(power = 0.8,
           sig.level = 0.05,
           d = 0.6)
```

```
     Two-sample t test power calculation

              n = 44.58577
              d = 0.6
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Power analysis in R: T-Test

```
library(pwr)

pwr.t.test(power = 0.8,
           sig.level = 0.05,
           d = 0.2)
```

```
     Two-sample t test power calculation

              n = 393.4057
              d = 0.2
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

A/B TESTING IN R

Let's practice!

A/B TESTING IN R

Statistical Tests

Page Piccinini
Instructor

Common statistical test for A/B testing

logistic regression - a binary (categorical) dependent variable (e.g., clicked or didn't click)

t-test (linear regression) - a continuous dependent variable (e.g., time spent on website)

T-tests

```
viz_website_2018_01 <- read_csv("viz_website_2018_01.csv")

aa_experiment_results <- t.test(
                                                          )
```

T-tests

```
viz_website_2018_01 <- read_csv("viz_website_2018_01.csv")

aa_experiment_results <- t.test(time_spent_homepage_sec
                                )
```

T-tests

```
viz_website_2018_01 <- read_csv("viz_website_2018_01.csv")

aa_experiment_results <- t.test(time_spent_homepage_sec ~ condition,
                                )
```

T-tests

```
viz_website_2018_01 <- read_csv("viz_website_2018_01.csv")

aa_experiment_results <- t.test(time_spent_homepage_sec ~ condition,
                                data = viz_website_2018_01)

aa_experiment_results
```

```
    Welch Two Sample t-test

data:  time_spent_homepage_sec by condition
t = -0.87836, df = 30998, p-value = 0.3798
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.03252741  0.01239578
sample estimates:
mean in group A1 mean in group A2
        58.99352         59.00358
```

T-test vs. linear regression

t-test (linear regression) - a continuous dependent variable (e.g., time spent on website)

T-test vs. linear regression

```
	Welch Two Sample t-test

data:  time_spent_homepage_sec by condition
t = -0.87836, df = 30998, p-value = 0.3798
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.03252741  0.01239578
sample estimates:
mean in group A1 mean in group A2
        58.99352         59.00358
```

```
lm(time_spent_homepage_sec ~ condition, data = viz_website_2018_01) %>%
  summary()
```

```
Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 58.993518   0.008103 7280.207   <2e-16 ***
conditionA2  0.010066   0.011460    0.878     0.38
```

A/B TESTING IN R

Let's practice!

A/B TESTING IN R

Stopping Rules and Sequential Analysis

Page Piccinini
Instructor

What is a stopping rule? - Cambridge Dictionary of Statistics

Stopping rules: Procedures that allow **interim analyses** in clinical trials at **predefined times**, while preserving the type I error at some pre-specified level.

What is a stopping rule? - Cambridge Dictionary of Statistics

Sequential analysis: A procedure in which a **statistical test** of significance is **conducted repeatedly** over time as the data are collected. After each observation, the cumulative data are analyzed and **one of the following** three decisions taken:

1. **stop** the data collection, reject the null hypothesis and claim **statistical significance**;
2. **stop** the data collection, do not reject the null hypothesis and state that the results are **not statistically significant**;
3. **continue** the data collection, since as yet the cumulated data are **inadequate to draw a conclusion**.

Why stopping rules are useful

- Prevent **p-hacking**.
- Accounts for **unsure effect size**.
- Allows for better **allocation of resources**.

Sequential analysis in R

```
library(gsDesign)

seq_analysis <- gsDesign(



                                        )
```

Sequential analysis in R

```
library(gsDesign)

seq_analysis <- gsDesign(k =
                         test.type =
                         alpha =
                         beta =
                         sfu =          )
```

Sequential analysis in R

```
library(gsDesign)

seq_analysis <- gsDesign(k = 4,
                         test.type =
                         alpha =
                         beta =
                         sfu =          )
```

Sequential analysis in R

```
library(gsDesign)

seq_analysis <- gsDesign(k = 4,
                         test.type = 1,
                         alpha =
                         beta =
                         sfu =           )
```

Sequential analysis in R

```
library(gsDesign)

seq_analysis <- gsDesign(k = 4,
                         test.type = 1,
                         alpha = 0.05,
                         beta =
                         sfu =          )
```

Sequential analysis in R

```
library(gsDesign)

seq_analysis <- gsDesign(k = 4,
                         test.type = 1,
                         alpha = 0.05,
                         beta = 0.2,
                         sfu =          )
```

Sequential analysis in R

```
library(gsDesign)

seq_analysis <- gsDesign(k = 4,
                         test.type = 1,
                         alpha = 0.05,
                         beta = 0.2,
                         sfu = "Pocock")

seq_analysis
```

```
One-sided group sequential design with
80 % power and 5 % Type I Error.
           Sample
            Size
  Analysis Ratio*  Z   Nominal p  Spend
         1  0.306 2.07    0.0193 0.0193
         2  0.612 2.07    0.0193 0.0132
         3  0.918 2.07    0.0193 0.0098
         4  1.224 2.07    0.0193 0.0077
     Total                       0.0500

++ alpha spending:
 Pocock boundary.
* Sample size ratio compared to fixed design with no interim
```

Sequential analysis in R

```
library(gsDesign)

seq_analysis <- gsDesign(k = 4,
                         test.type = 1,
                         alpha = 0.05,
                         beta = 0.2,
                         sfu = "Pocock")
seq_analysis

max_n <- 1000

max_n_per_group <- max_n / 2

stopping_points <- max_n_per_group * seq_analysis$timing
stopping_points
```

```
[1] 125 250 375 500
```

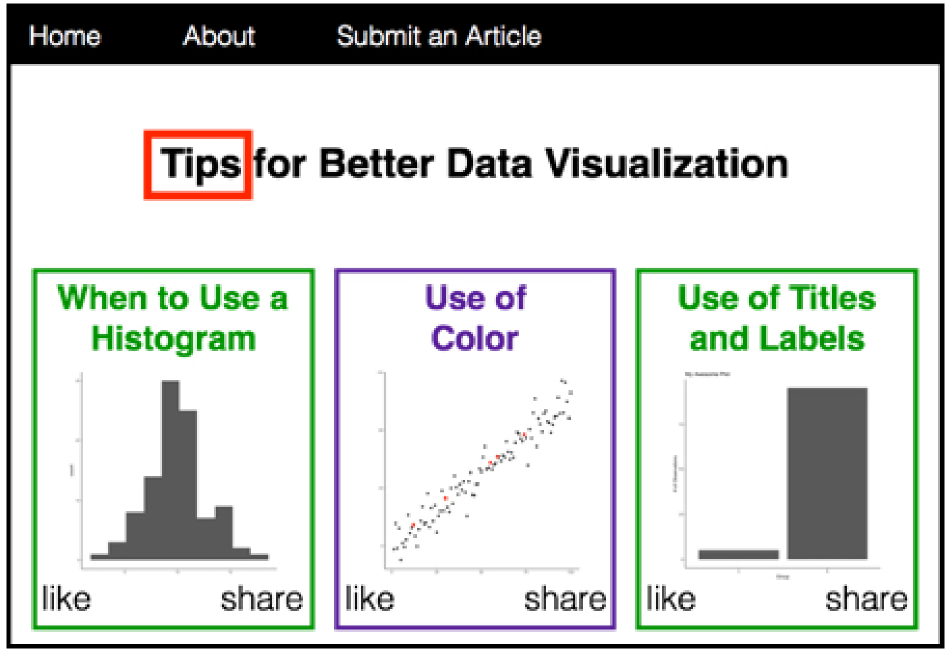
A/B TESTING IN R

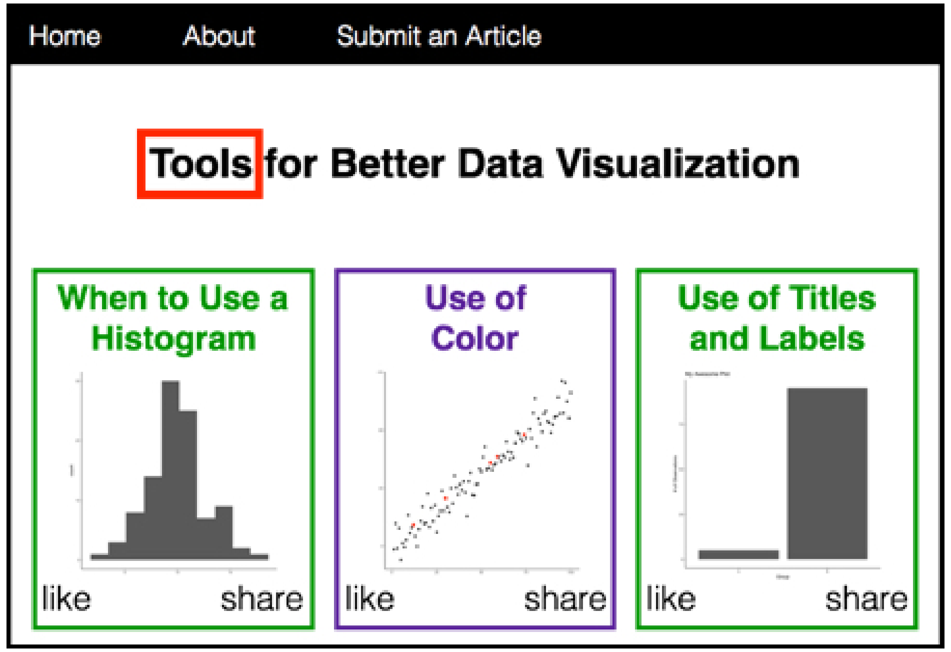
Let's practice!

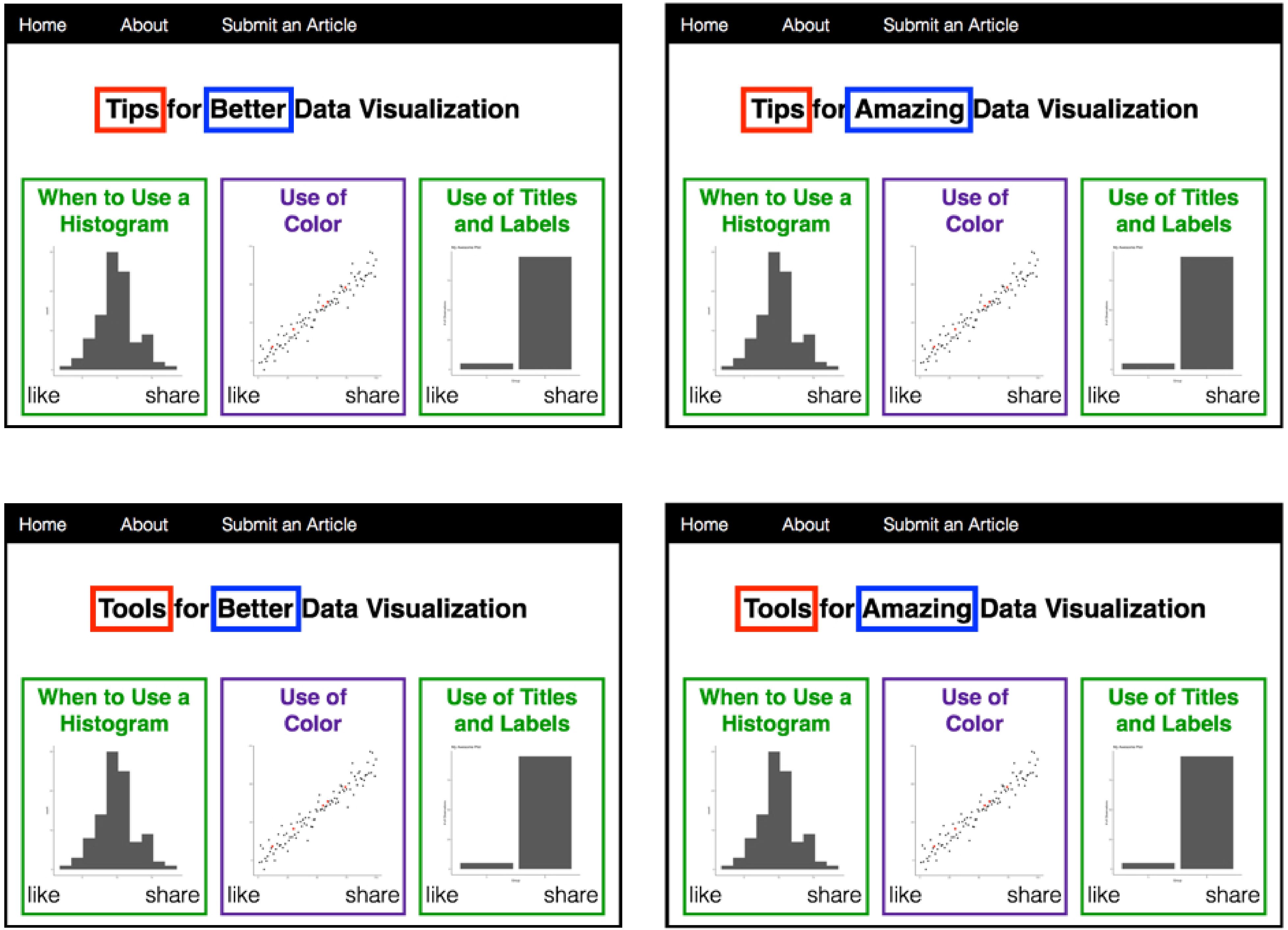
A/B TESTING IN R

Multivariate Testing

Page Piccinini
Instructor


Home
About
Submit an Article
Tips for Better Data Visualization
When to Use a Histogram
like
share
Use of Color
like
share
Use of Titles and Labels
like
share


Home
About
Submit an Article
Tools for Better Data Visualization
When to Use a Histogram
like
share
Use of Color
like
share
Use of Titles and Labels
like
share


Home
About
Submit an Article
Tips for Better Data Visualization
When to Use a Histogram
Use of Color
Use of Titles and Labels
like
share
like
share
like
share
Home
About
Submit an Article
Tips for Amazing Data Visualization
When to Use a Histogram
Use of Color
Use of Titles and Labels
like
share
like
share
like
share
Home
About
Submit an Article
Tools for Better Data Visualization
When to Use a Histogram
Use of Color
Use of Titles and Labels
like
share
like
share
like
share
Home
About
Submit an Article
Tools for Amazing Data Visualization
When to Use a Histogram
Use of Color
Use of Titles and Labels
like
share
like
share
like
share

Time spent on homepage multivariate analysis

```
library(broom)

multivar_results <- lm(time_spent_homepage_sec ~
                         data = viz_website_2018_05) %>%
  tidy()
```

Time spent on homepage multivariate analysis

```
library(broom)

multivar_results <- lm(time_spent_homepage_sec ~ word_one
                       data = viz_website_2018_05) %>%
  tidy()
```

Time spent on homepage multivariate analysis

```
library(broom)

multivar_results <- lm(time_spent_homepage_sec ~ word_one * word_two,
                       data = viz_website_2018_05) %>%
  tidy()

multivar_results
```

```
                            term    estimate   std.error  statistic   p.value
1                    (Intercept) 48.00829170 0.008056696 5958.80671 0.0000000
2                  word_onetools  4.98549854 0.011393888  437.55902 0.0000000
3                 word_twobetter -0.01323206 0.011393888   -1.16133 0.2455122
4 word_onetools:word_twobetter -4.97918356 0.016113391 -309.00904 0.0000000
```

Time spent on homepage multivariate analysis

```
library(broom)

multivar_results <- viz_website_2018_05 %>%
  mutate(word_one = factor(word_one,
                           levels = c("tips", "tools"))) %>%
  mutate(word_two = factor(word_two,
                           levels = c("better", "amazing")))
```

Time spent on homepage multivariate analysis

```
library(broom)

multivar_results <- viz_website_2018_05 %>%
  mutate(word_one = factor(word_one,
                           levels = c("tips", "tools"))) %>%
  mutate(word_two = factor(word_two,
                           levels = c("better", "amazing"))) %>%
  lm(time_spent_homepage_sec ~ word_one * word_two,
     data = .) %>%
  tidy()

multivar_results
```

```
                            term    estimate   std.error    statistic   p.value
1                    (Intercept) 47.995059637 0.008056696 5957.1643430 0.0000000
2                  word_onetools  0.006314972 0.011393888    0.5542421 0.5794152
3                word_twoamazing  0.013232063 0.011393888    1.1613299 0.2455122
4 word_onetools:word_twoamazing  4.979183565 0.016113391  309.0090419 0.0000000
```

A/B TESTING IN R

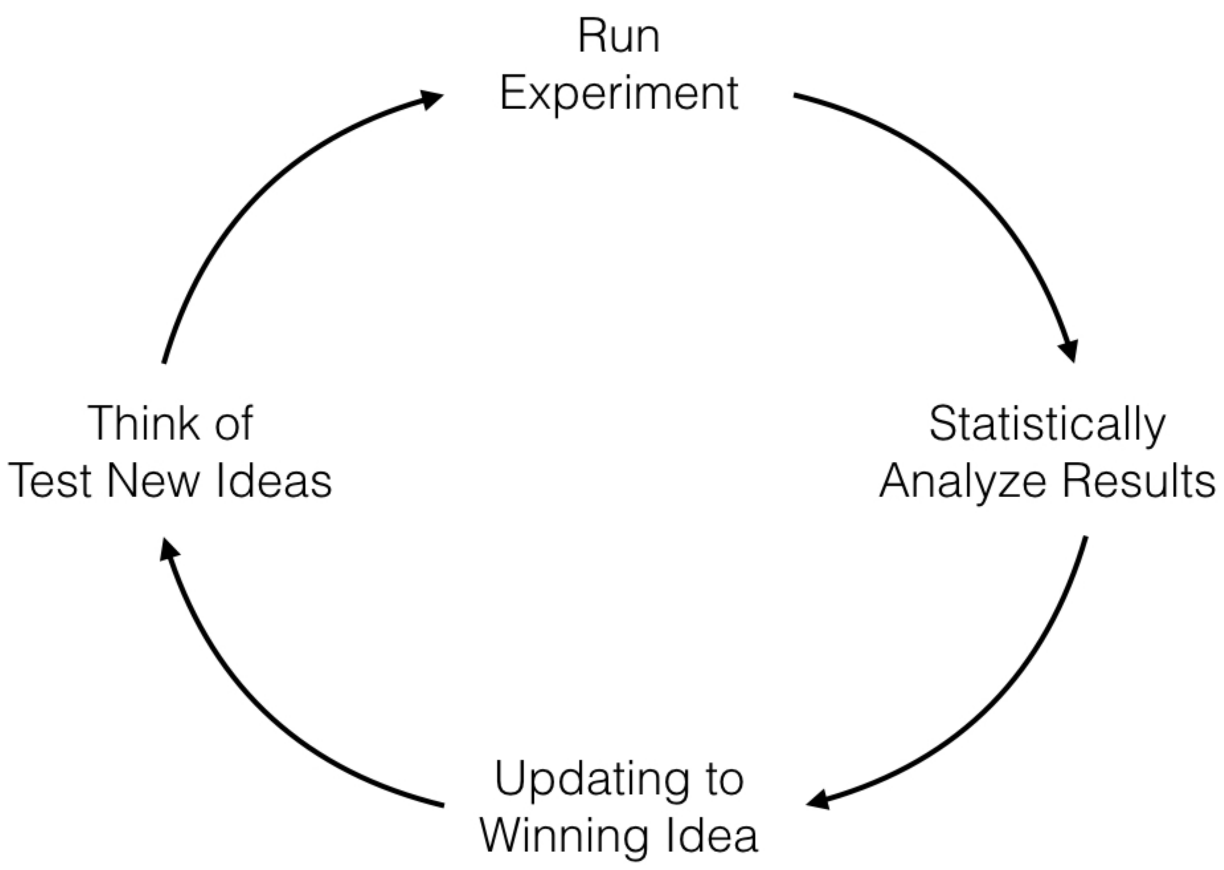
Let's practice!

A/B TESTING IN R

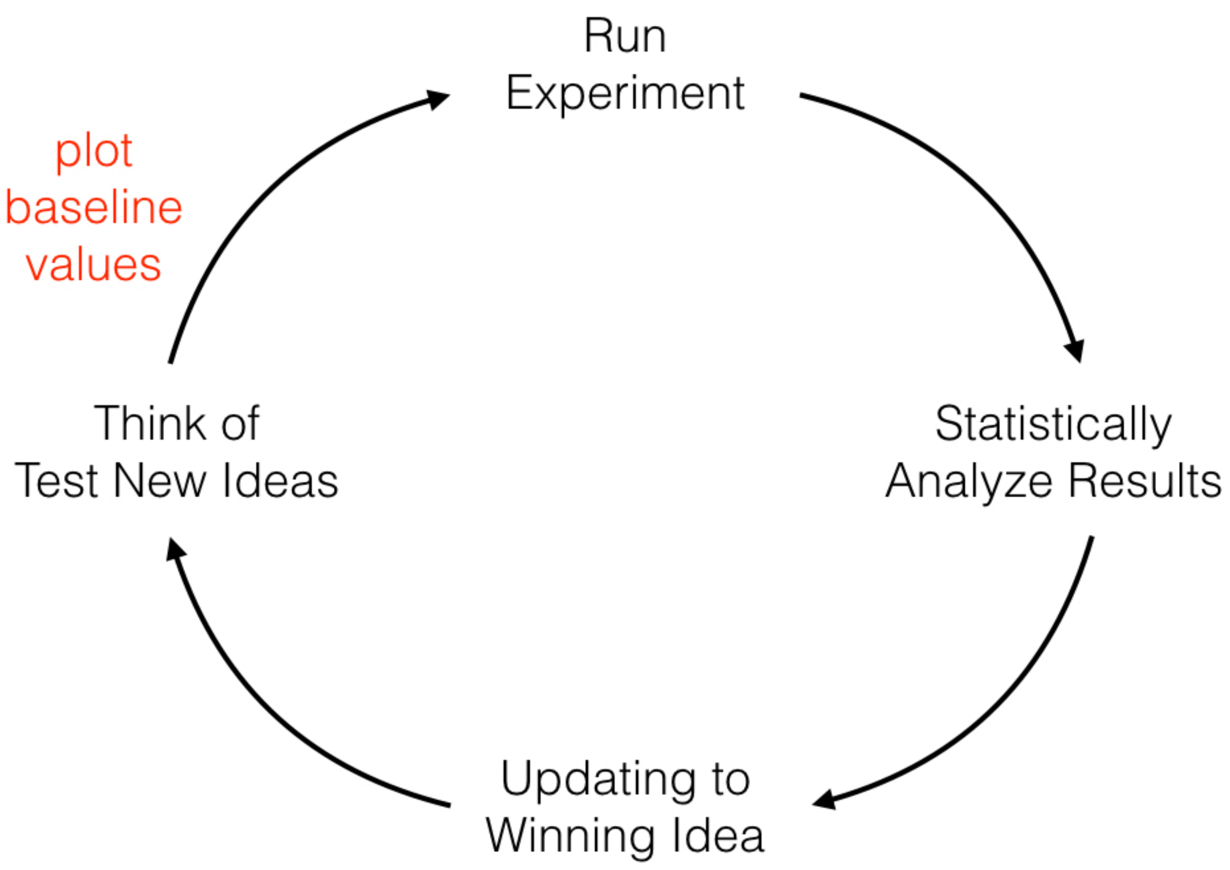
A/B Testing Recap

Page Piccinini
Instrutor

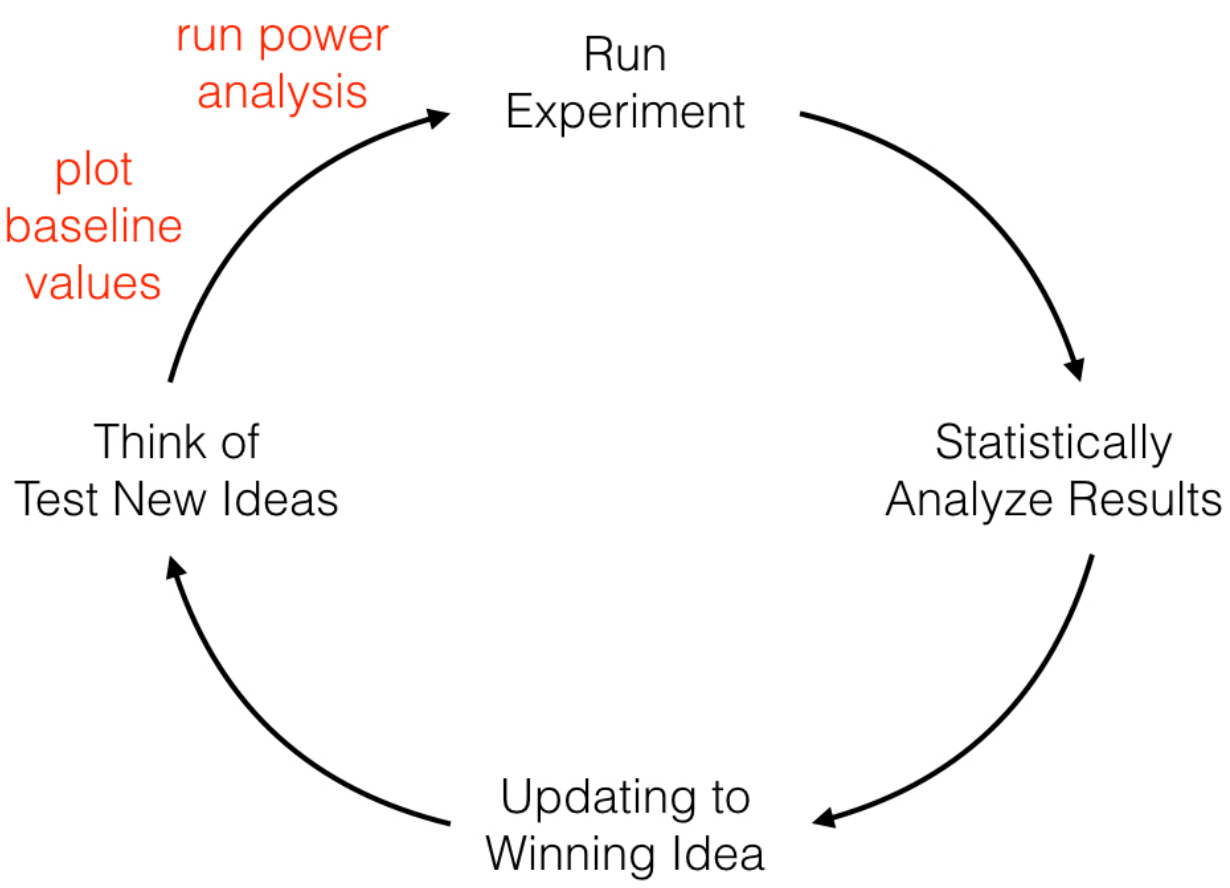
A/B testing summary



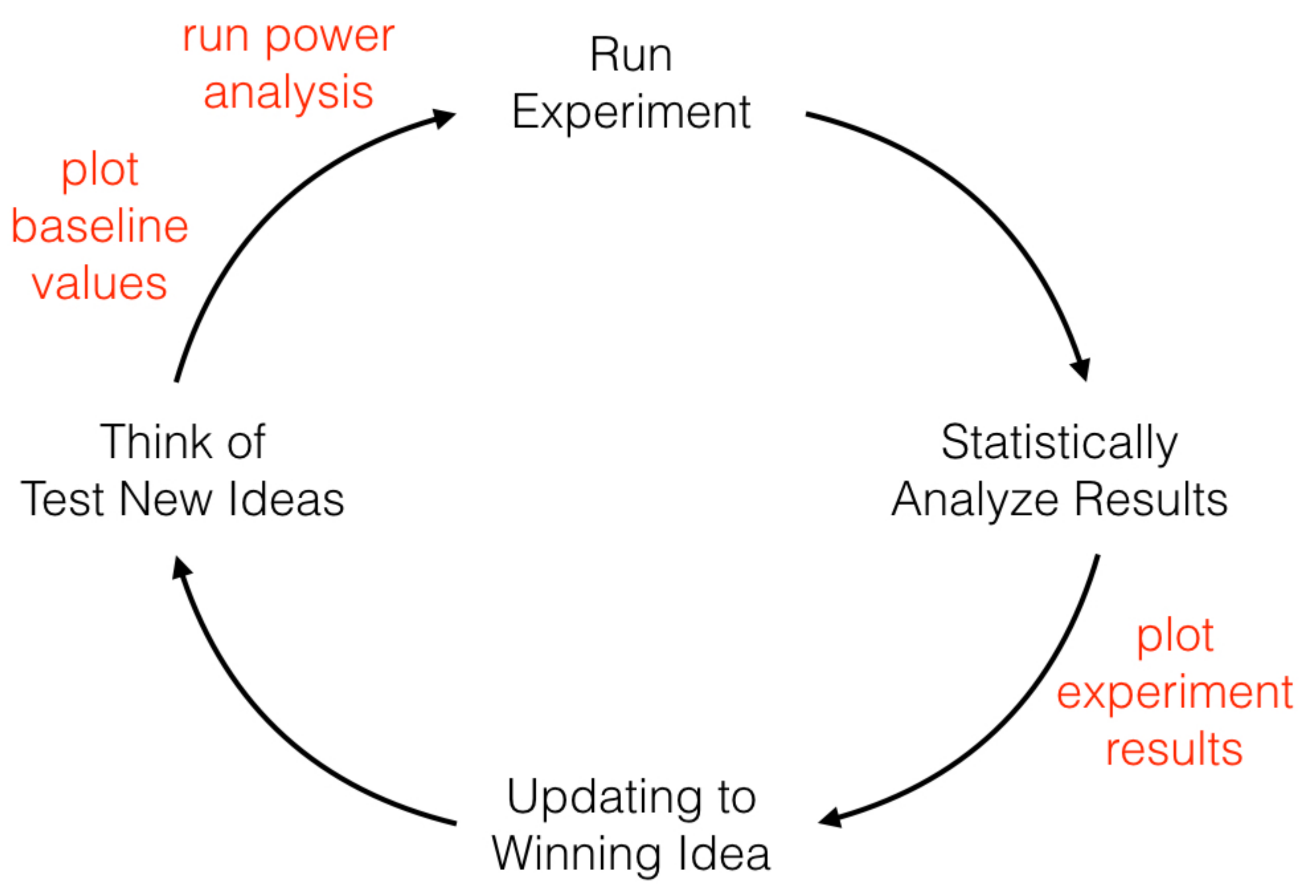
A/B testing summary



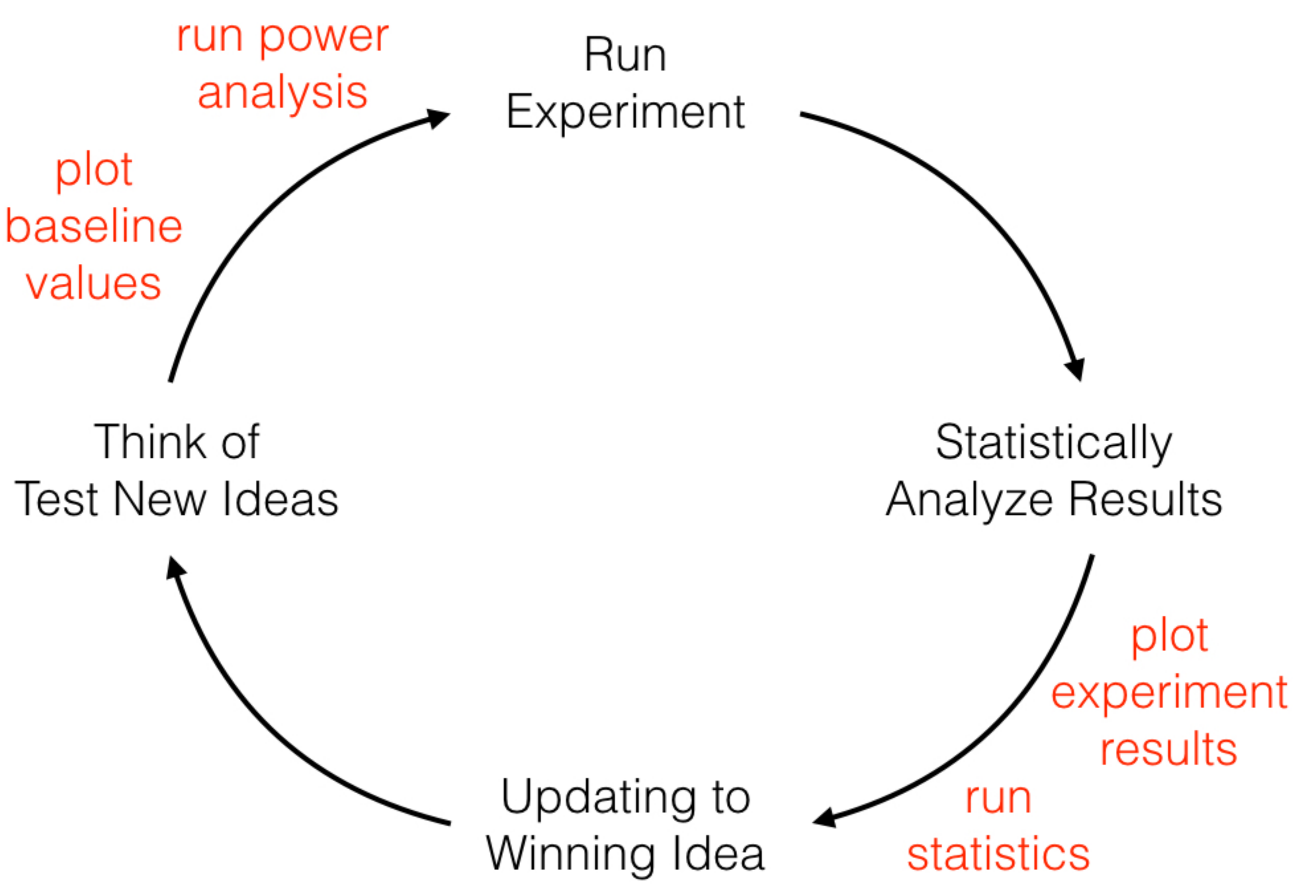
A/B testing summary



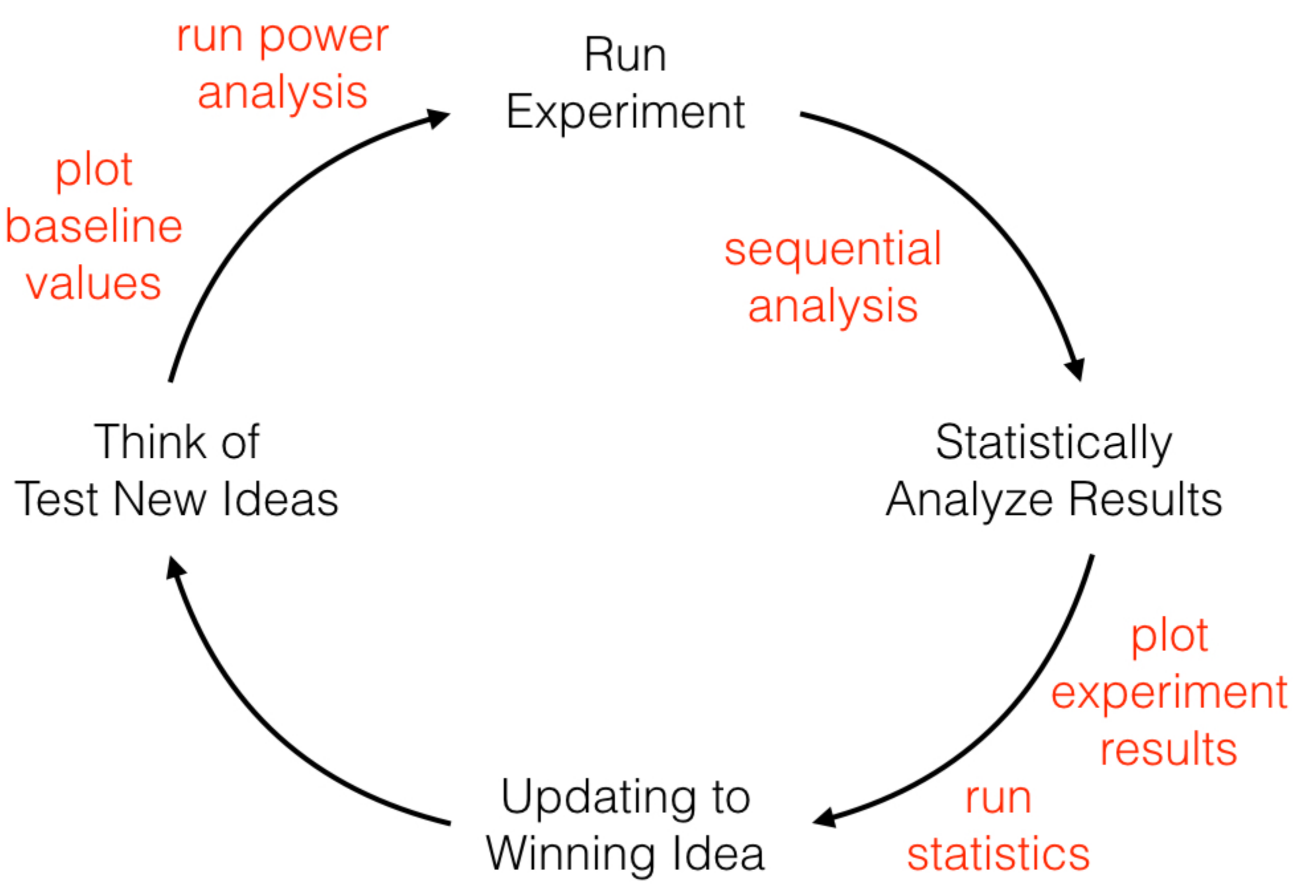
A/B testing summary



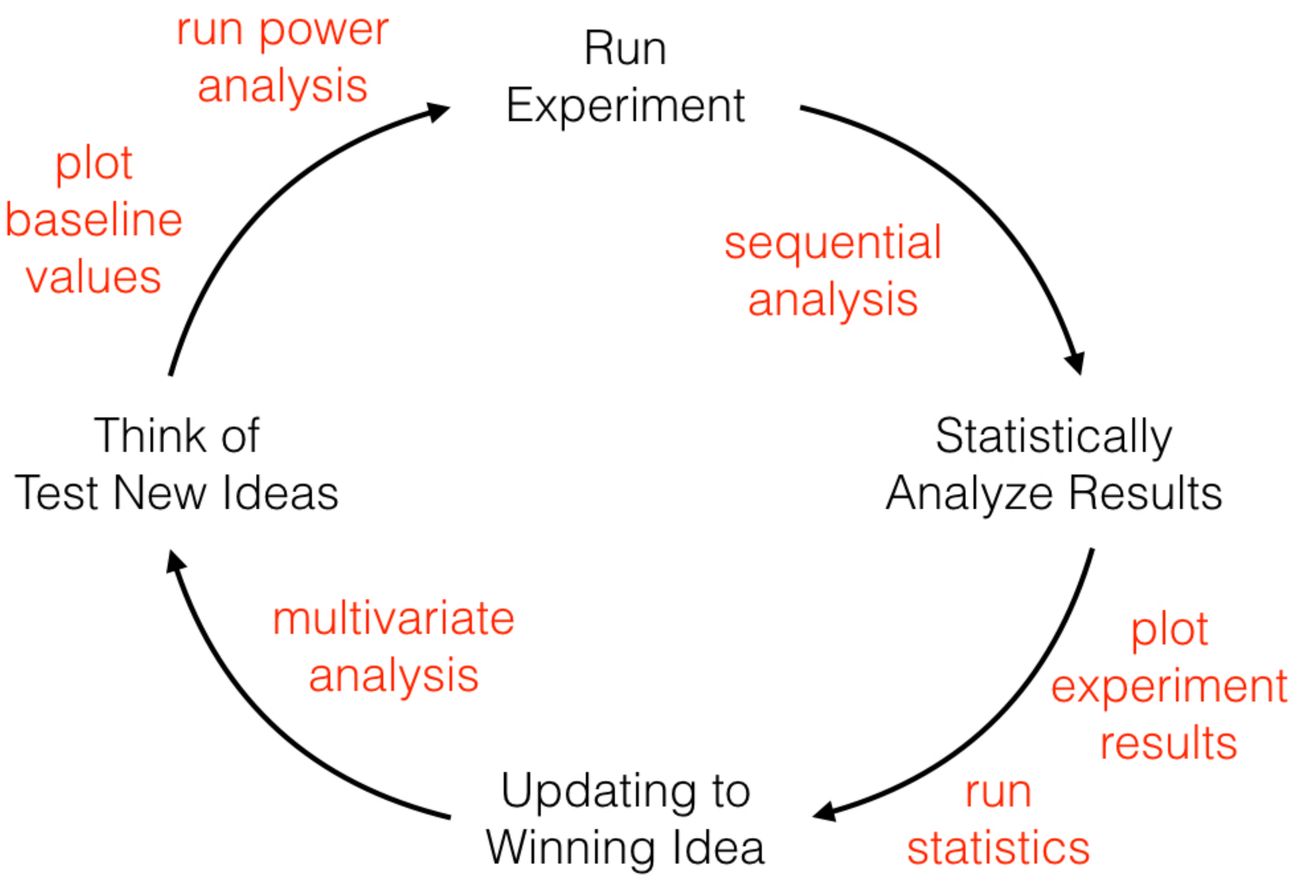
A/B testing summary



A/B testing summary



A/B testing summary



A/B TESTING IN R

Thank you!