

Data Visualization

KickStarter Projects 2018

John Liao

CS 544 Fall 2018 - Foundations of Analytics with R

Boston University Metropolitan College

Picking the Data Set

- Selected Kickstarter projects from 2018 data set¹
- More than 300,000 kickstarter projects
- Data Set Variables include:
 - Main Category
 - Sub Category
 - Deadline date
 - Launch Date
 - Goal amounts (\$)
 - Pledged amounts (\$)
 - Country
 - Number of backers
 - State (success/failure/cancelled)

KICKSTARTER



This is a card game for people who are into kittens and explosions and laser beams and sometimes goats.

[Buy it now!](#)

Created by
Elan Lee

219,382 backers pledged \$8,782,571 to help bring this project to life.

Preparing the Data

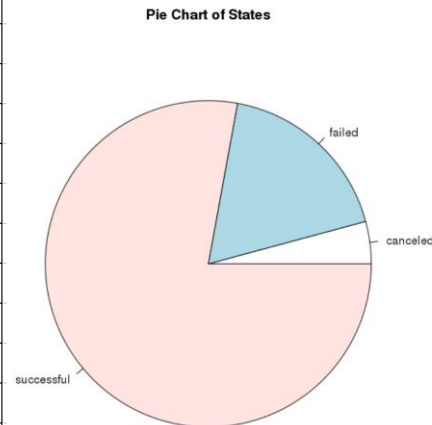
- Limit scope: remove KickStarters where
 - Pledged < \$1000
 - # Backers <= 25
 - State is **live** or **suspended**
 - Country is not US
- Scope of analysis: compare **regular** vs. **unicorn** KickStarters
 - **Regular**: >\$1K and <\$1M Pledged (USD)
 - **Unicorn**: >\$1M Pledged (USD)

Analysis of Categorical Variable

Regular

- **103443** kick starter projects
- **159** unique sub categories

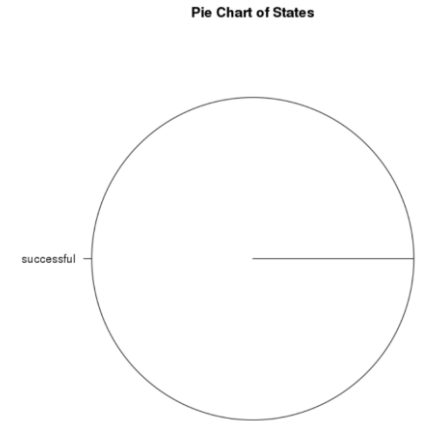
Top 10 Breakdown of Sub Categories		
Regular Kickstarters		
Category	% of Sample Size	# Projects
Product Design	7.8%	8069
Tabletop Games	6.7%	6931
Documentary	5.7%	5896
Music	4.5%	4655
Shorts	4.0%	4138
Food	3.5%	3621
Theater	3.0%	3103
Indie Rock	2.8%	2896
Video Games	2.6%	2690
Rock	2.5%	2586



Unicorn

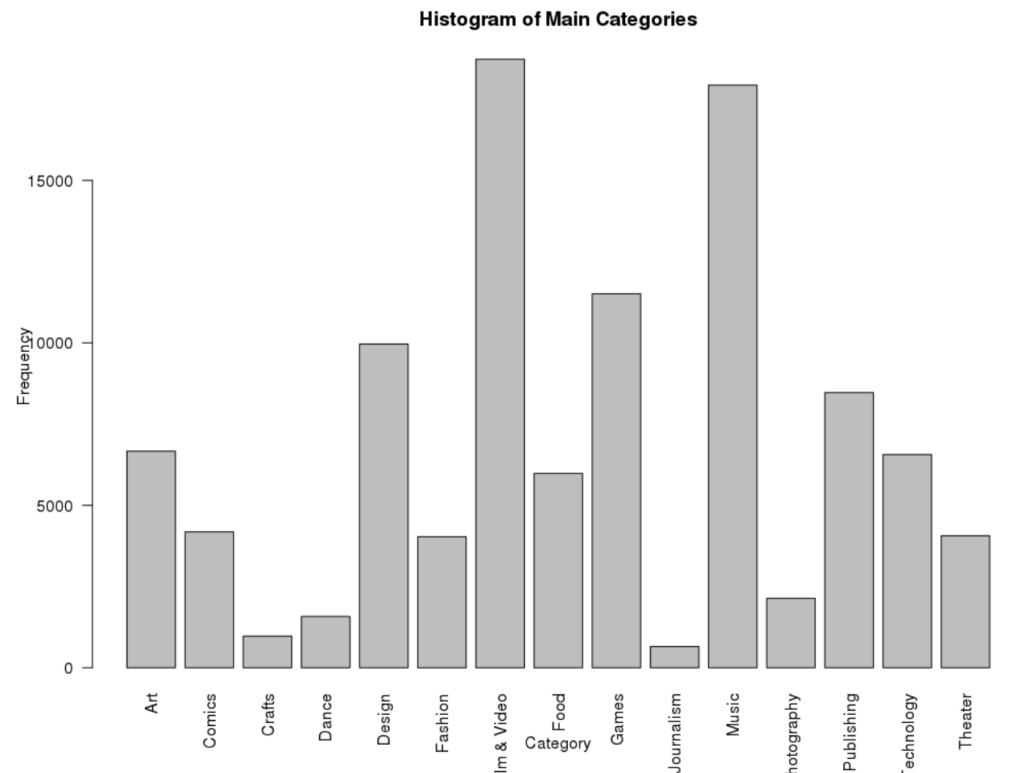
- **175** kick starter projects
- **30** unique sub categories

Top 10 Breakdown of Sub Categories		
Unicorn Kickstarters		
Category	% of Sample Size	# Projects
Tabletop Games	22.9%	40
Product Design	16.0%	28
Video Games	16.0%	28
Hardware	8.0%	14
Technology	6.9%	12
3D Printing	2.9%	5
Food	2.9%	5
Gadgets	2.9%	5
Design	2.3%	4
Narrative Film	2.3%	4

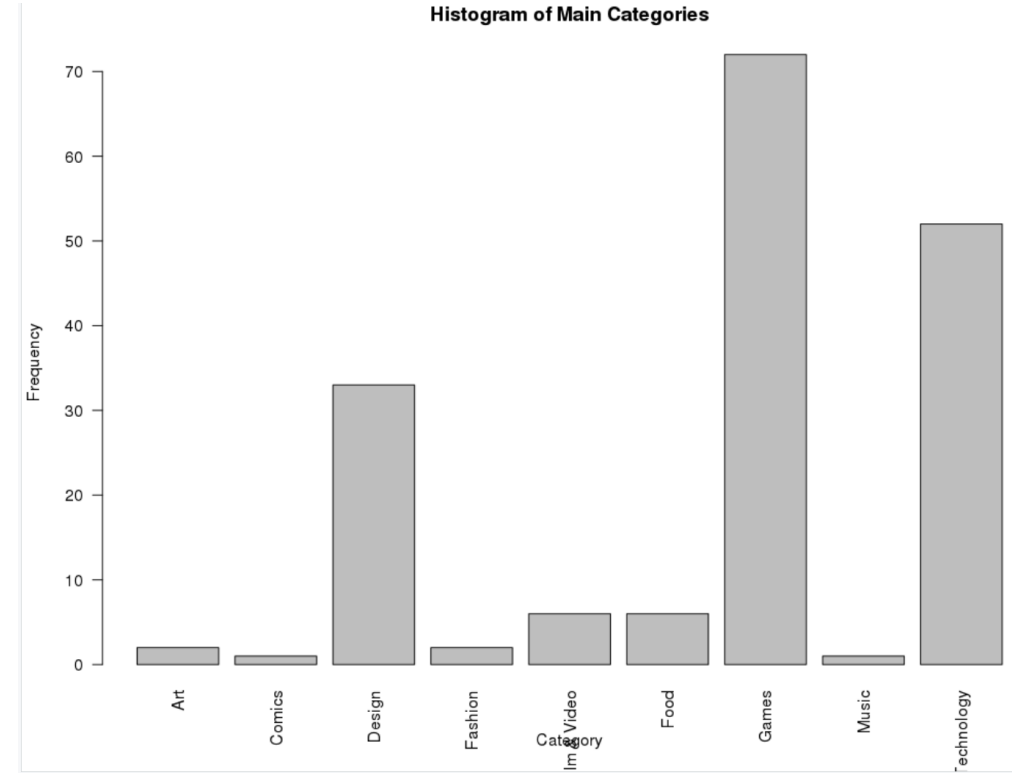


Analysis of Categorical Variable

Regular



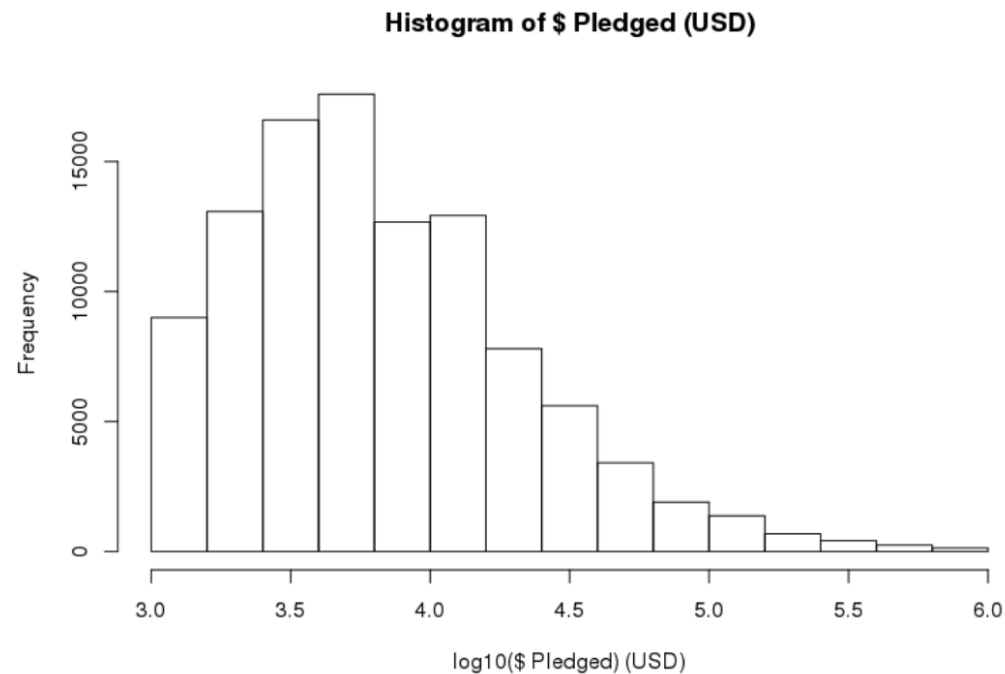
Unicorn



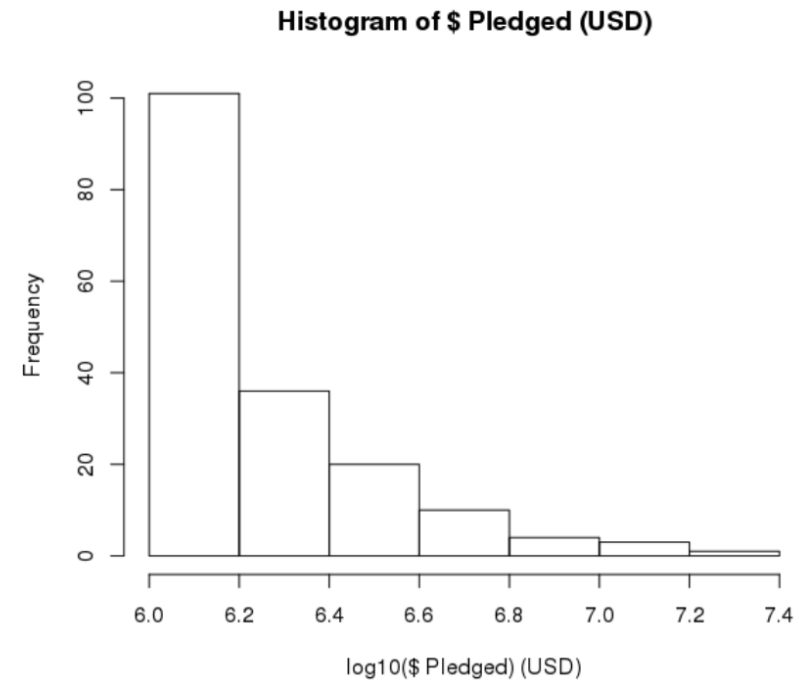
- For Regular KickStarters, **Music and Film & Video** is the top main category
- For Unicorn KickStarters, **Games and Technology** are the top main category selections
- For both Unicorn and Regular KickStarters, **Games** is a popular project

Analysis of Numerical Variable

Regular



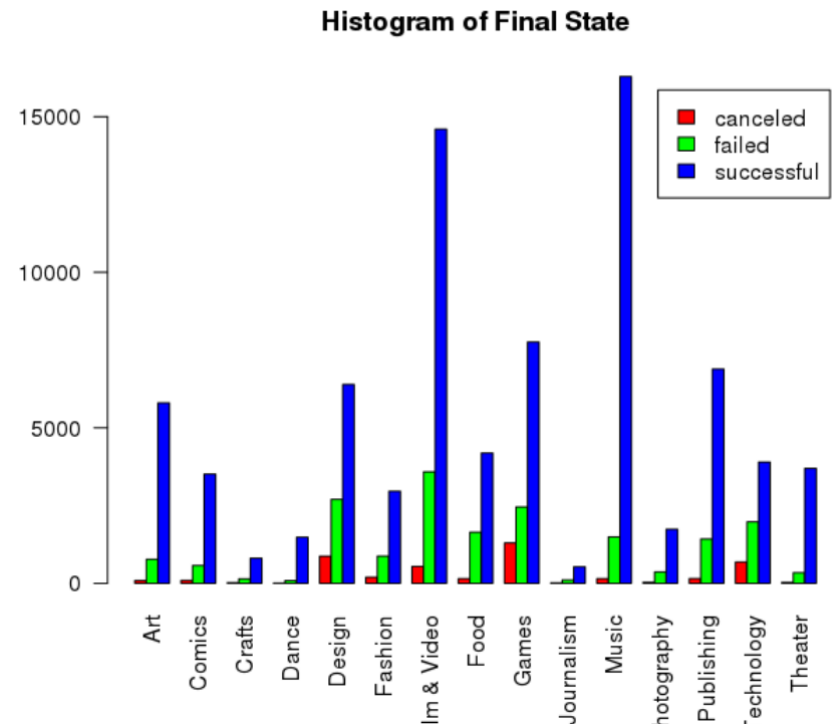
Unicorn



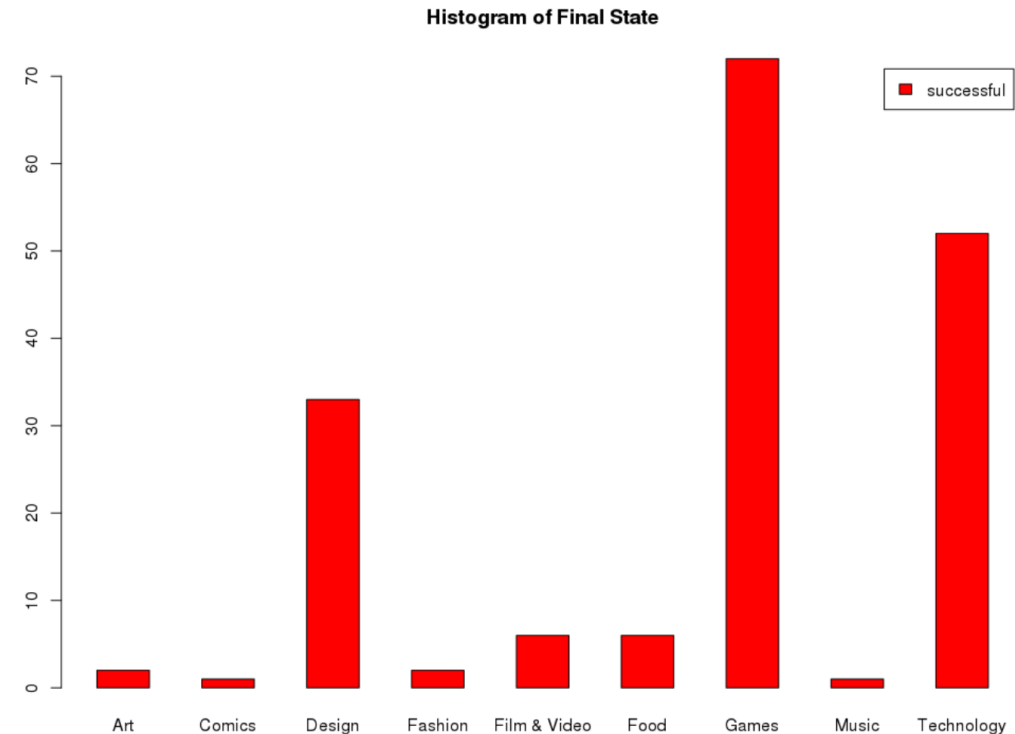
- Across Regular and Unicorn KickStarters, pledged amount follows a left-skewed distribution
- Distribution peaks around $10^{3.6}$ (Approx. \$4K)

Analysis of Set of Two or More Variables

Regular



Unicorn



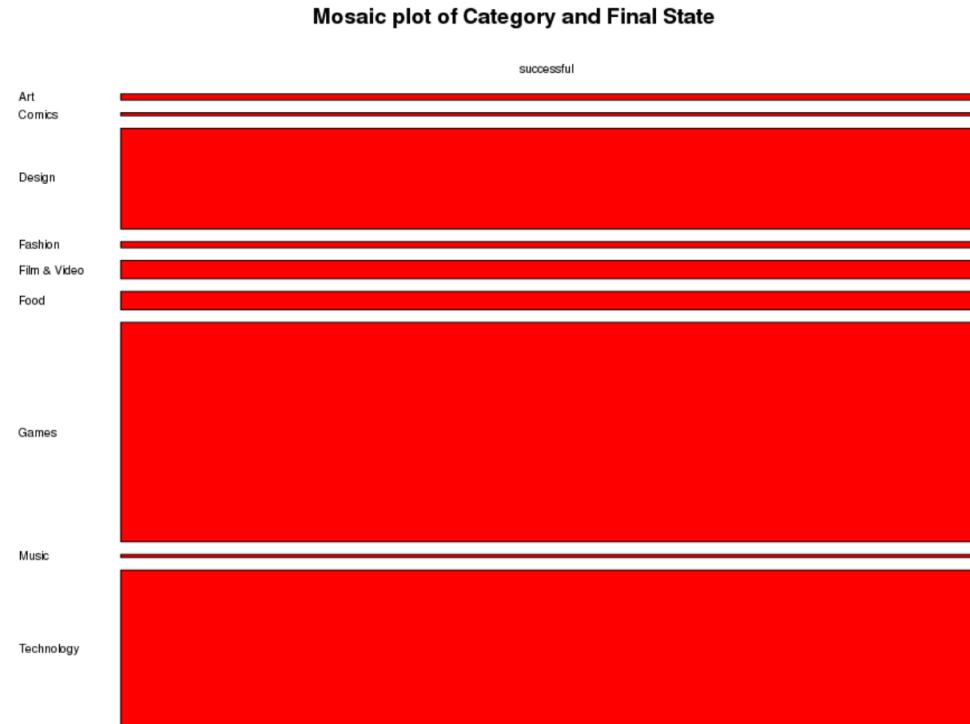
- For Regular KickStarters, **Music** was the most successful and **Games** failed the most
- For Unicorn Kickstarters, all projects were successful.
- For Unicorn Kickstarters, **Games** were the most plentiful

Analysis of Set of Two or More Variables

Regular



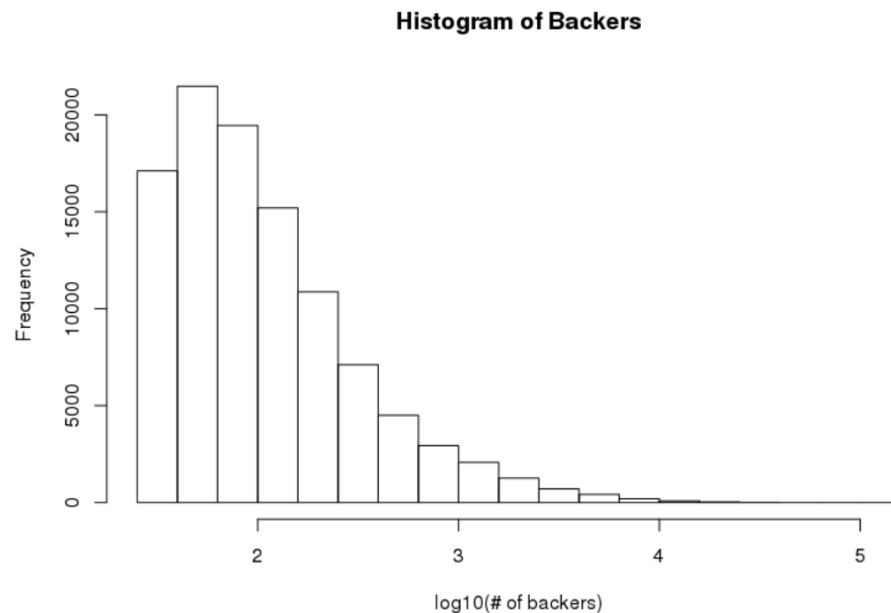
Unicorn



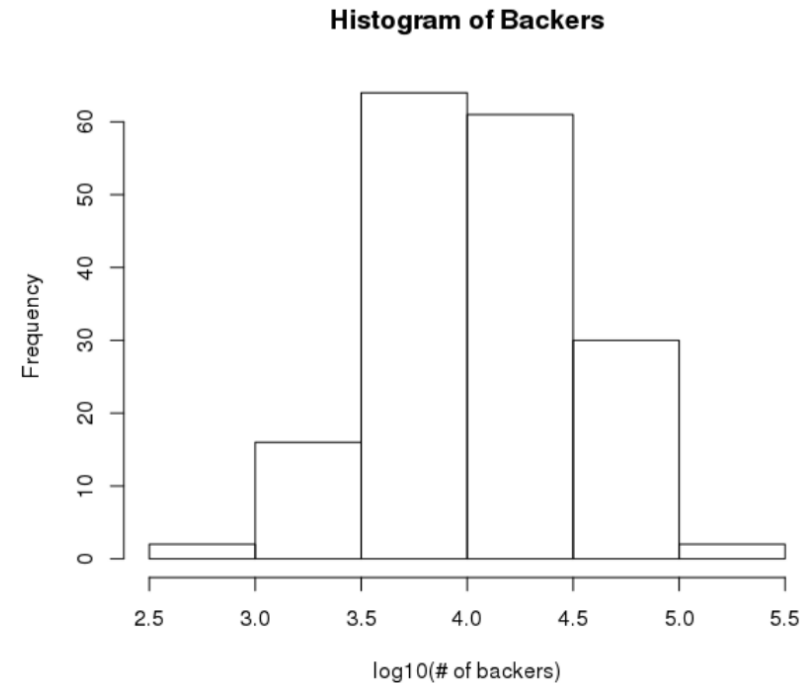
- For Regular KickStarters, **Technology and Design** has the highest failure and cancelled rates
- For Regular KickStarters, **Food** appears to have the highest fail rate for such a small number of projects
- For Regular KickStarters, **Art and Theater** appear to have a high success rate for such a small number of projects

Analysis of Numerical Variable and Examine Distribution

Regular



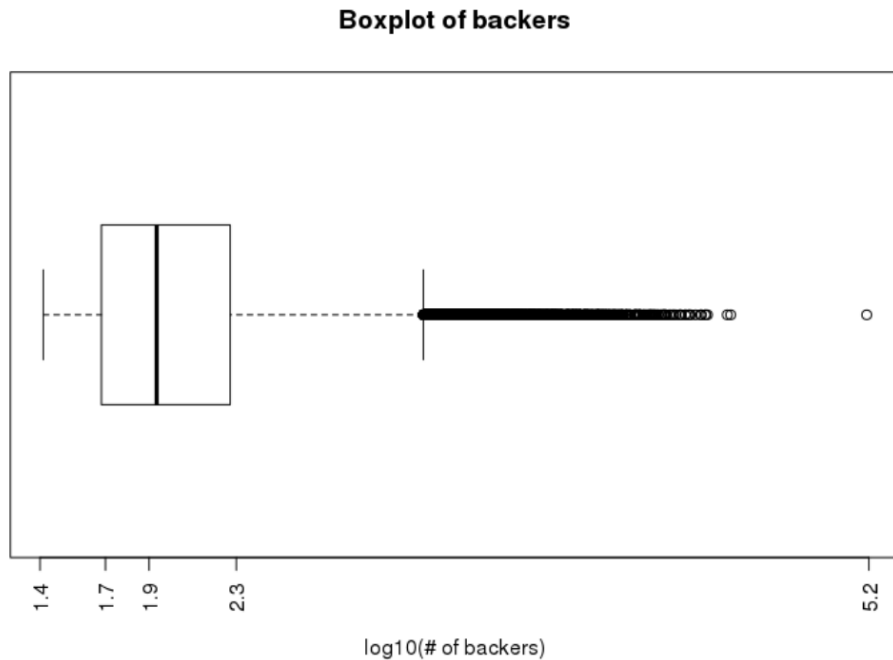
Unicorn



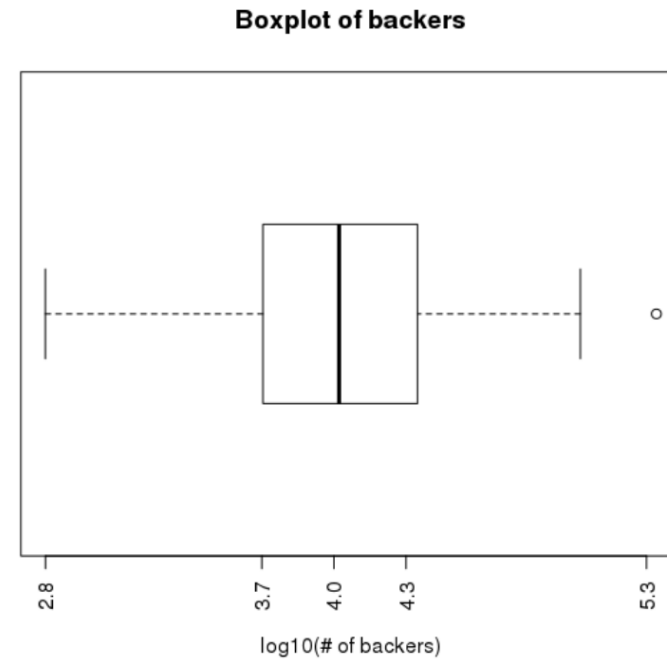
- Left skewed distribution with peak around $10^{1.5}$ (30) backers for Regular KickStarters
- Normal distribution with peak around 10^4 (10,000) backers for Unicorn KickStarters

Analysis of Numerical Variable and Examine Distribution

Regular



Unicorn



- For Regular KickStarters, a significant amount of projects are outliers

Show Applicability of Central Limit Theorem

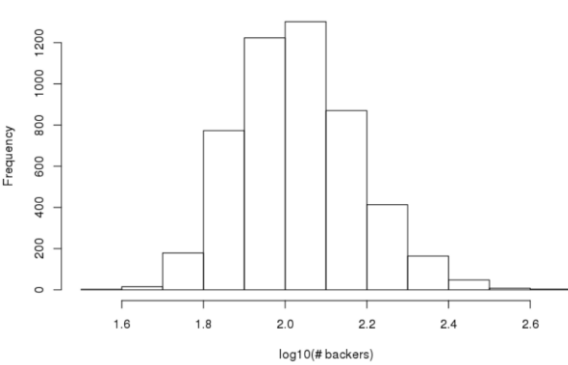
Regular

$$\mu = 2.36$$
$$\sigma = 0.47$$

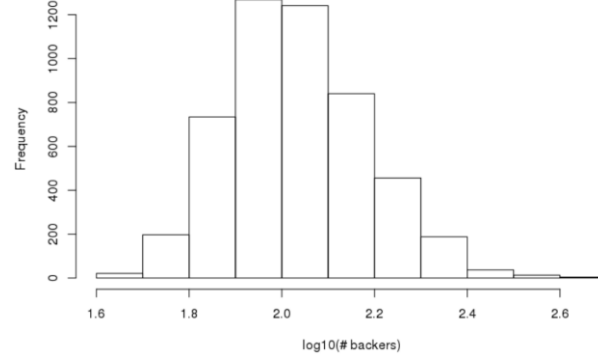
Unicorn

$$\mu = 4.04$$
$$\sigma = 0.47$$

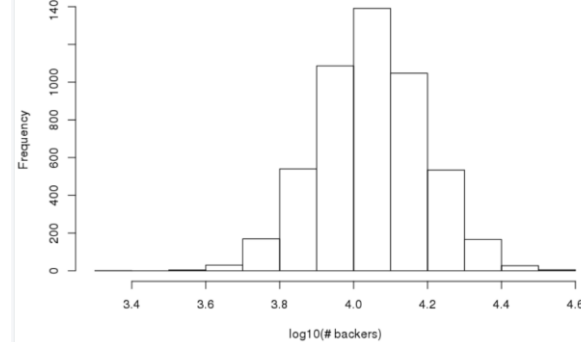
Histogram of densities of sample mean of # backers
SAMPLES=500, SAMPLE SIZE=10



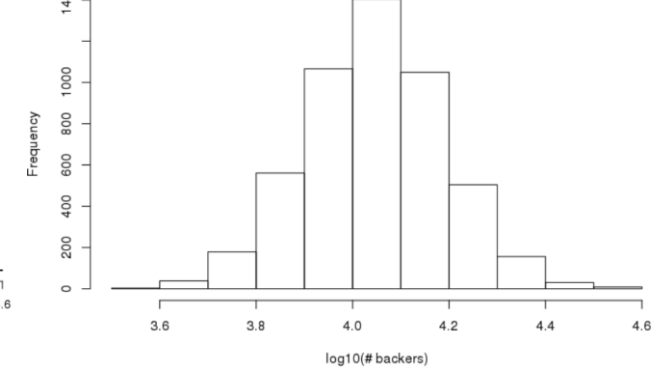
Histogram of densities of sample mean of # backers
SAMPLES=50, SAMPLE SIZE=10



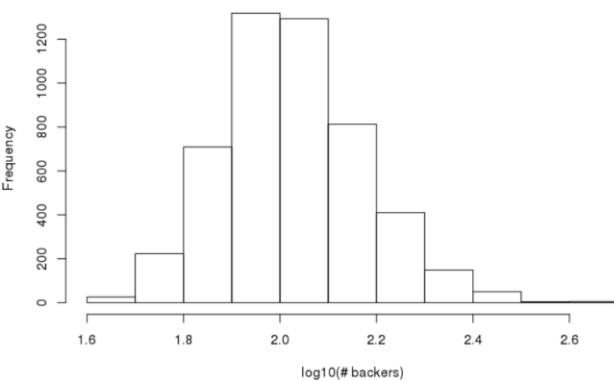
Histogram of densities of sample mean of # backers
SAMPLES=5, SAMPLE SIZE=10



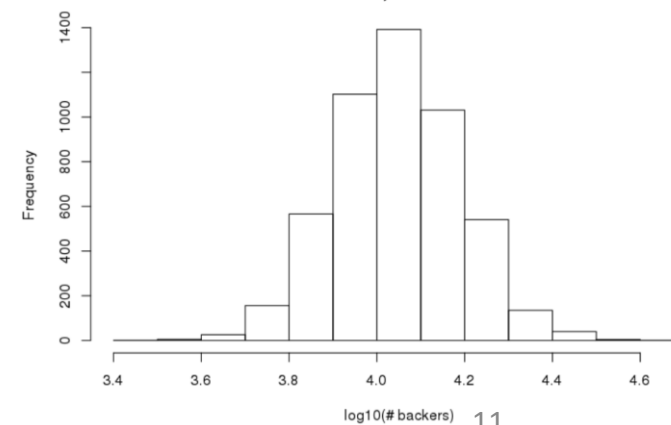
Histogram of densities of sample mean of # backers
SAMPLES=50, SAMPLE SIZE=10



Histogram of densities of sample mean of # backers
SAMPLES=5, SAMPLE SIZE=10



Histogram of densities of sample mean of # backers
SAMPLES=500, SAMPLE SIZE=10



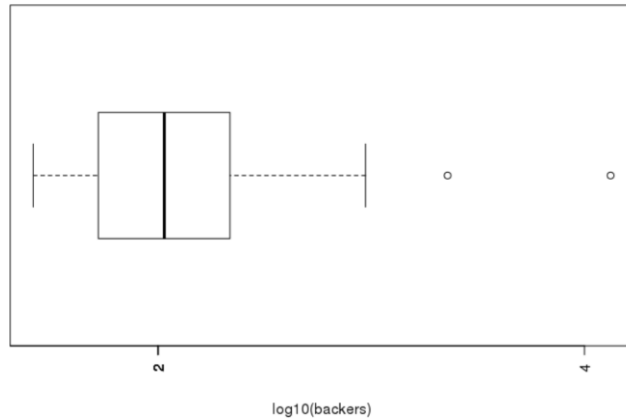
- As sample size goes up, the mean of sample means gets closer to the true mean and the sd of sample means decreases
- Increase in the sample sizes result in the distribution becoming less skewed and slowly approaching the shape of a normal distribution
- Samples drawn display a normal distribution

Draw Samples Using Various Sampling Methods

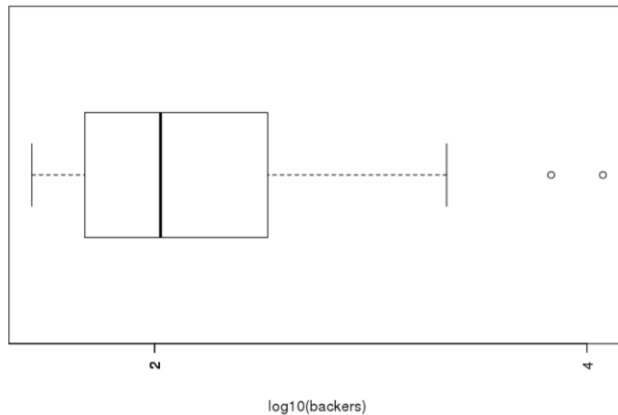
Regular

$$\mu = 2.36$$
$$\sigma = 0.47$$

Sampling Method #1: SRSWOR For # Backers



Sampling Method #2 For # Backers: Systematic Sampling

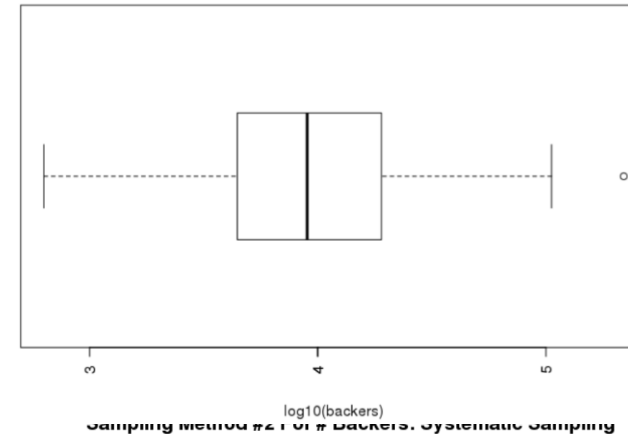


SRSWOR
100 samples drawn

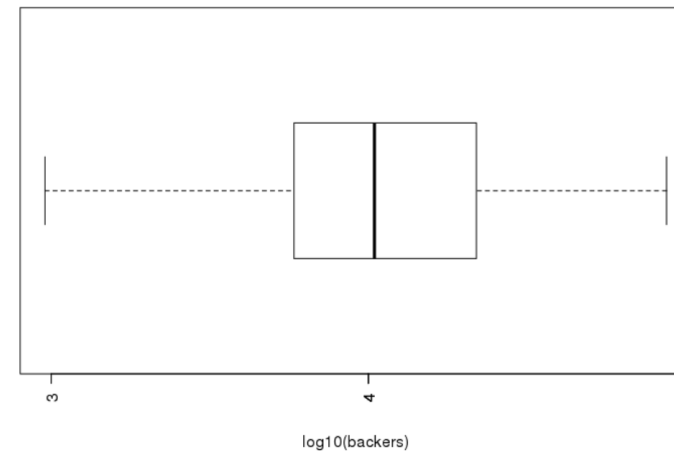
Unicorn

$$\mu = 4.04$$
$$\sigma = 0.47$$

Sampling Method #1: SRSWOR For # Backers



Sampling Method #2 For # Backers: Systematic Sampling



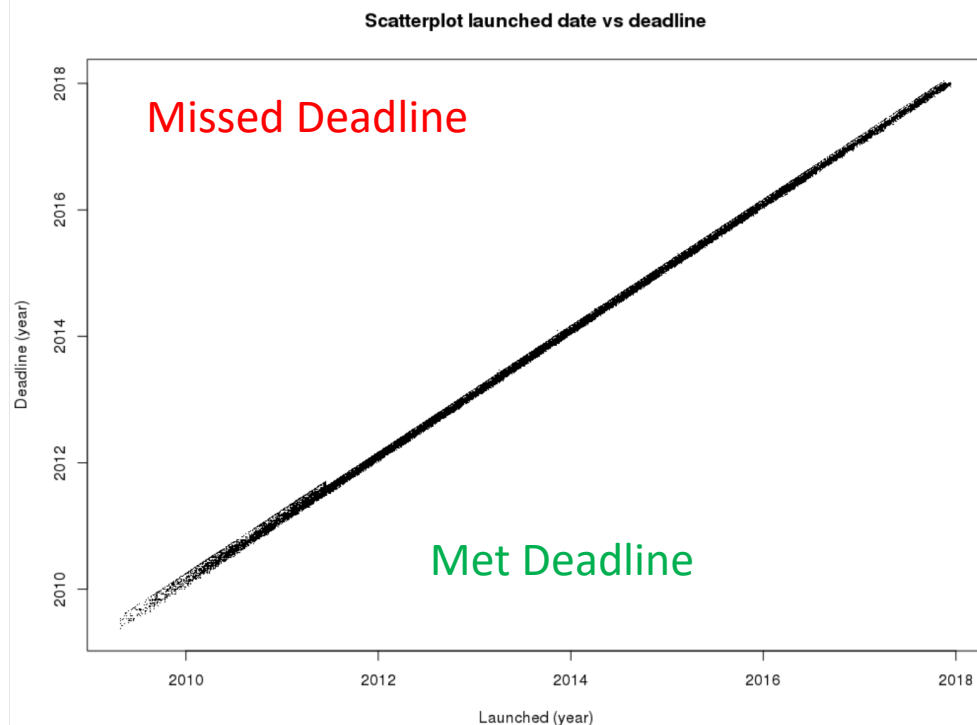
Systematic Sampling
Sample every
100th item

Conclusions if Samples Are Used Instead of Whole Dataset

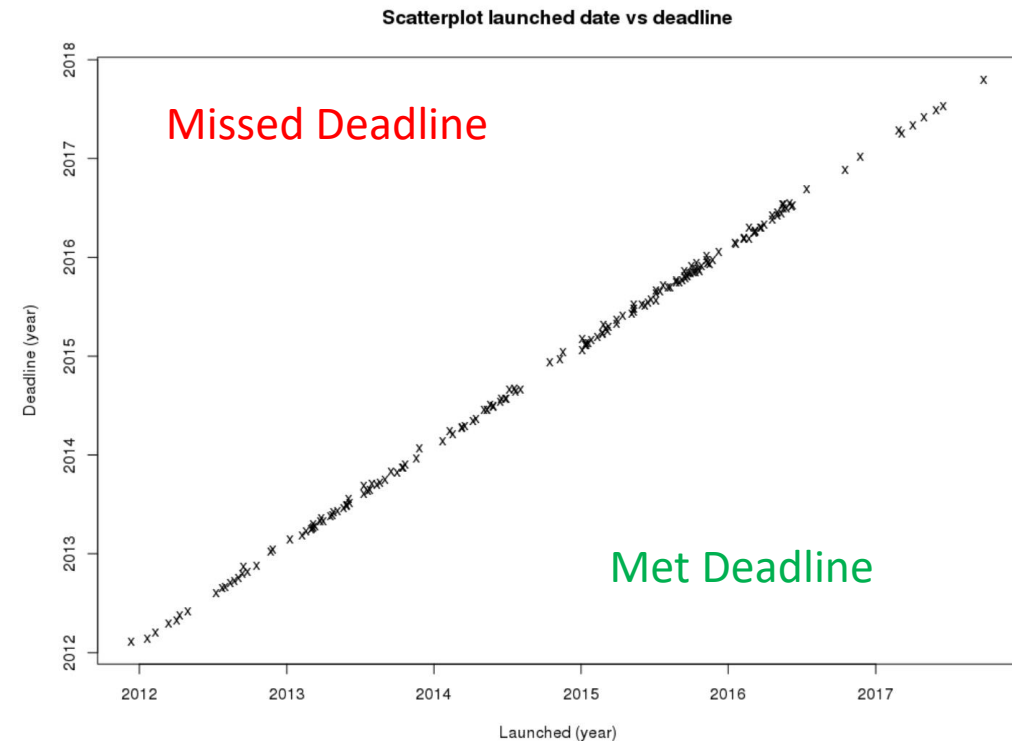
- Simple Random Sampling Without Replacement (SRSWOR) is free from bias
- Systematic Sampling subject to skew if the order of the items is not random

Implementation of Features Not Mentioned in Specification

Regular



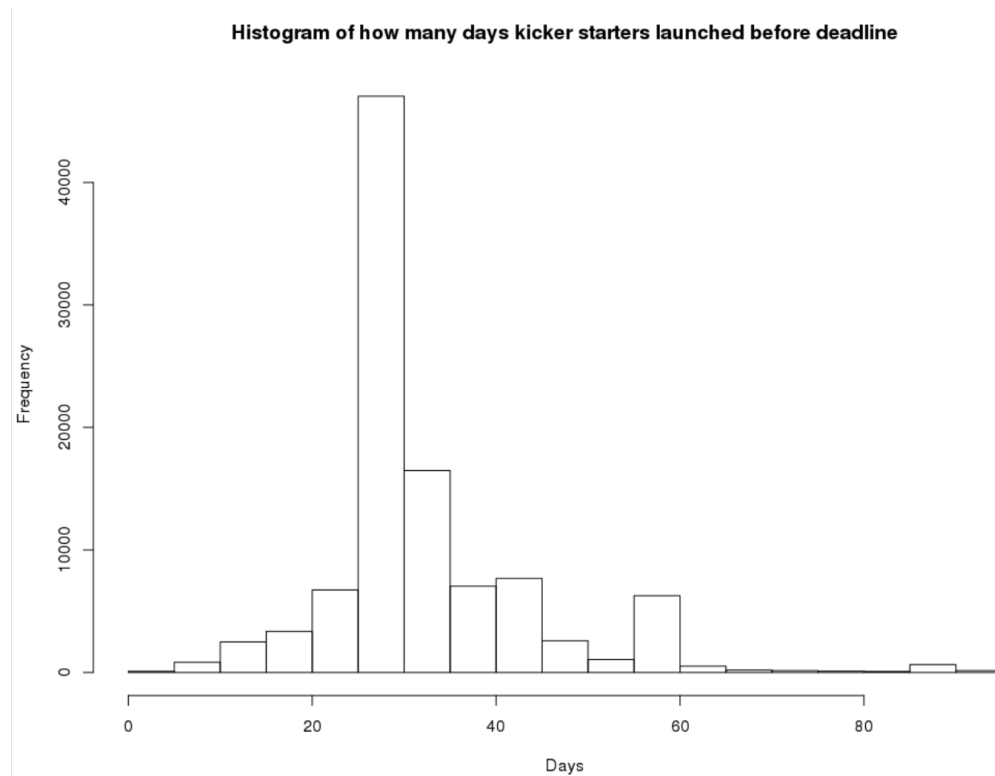
Unicorn



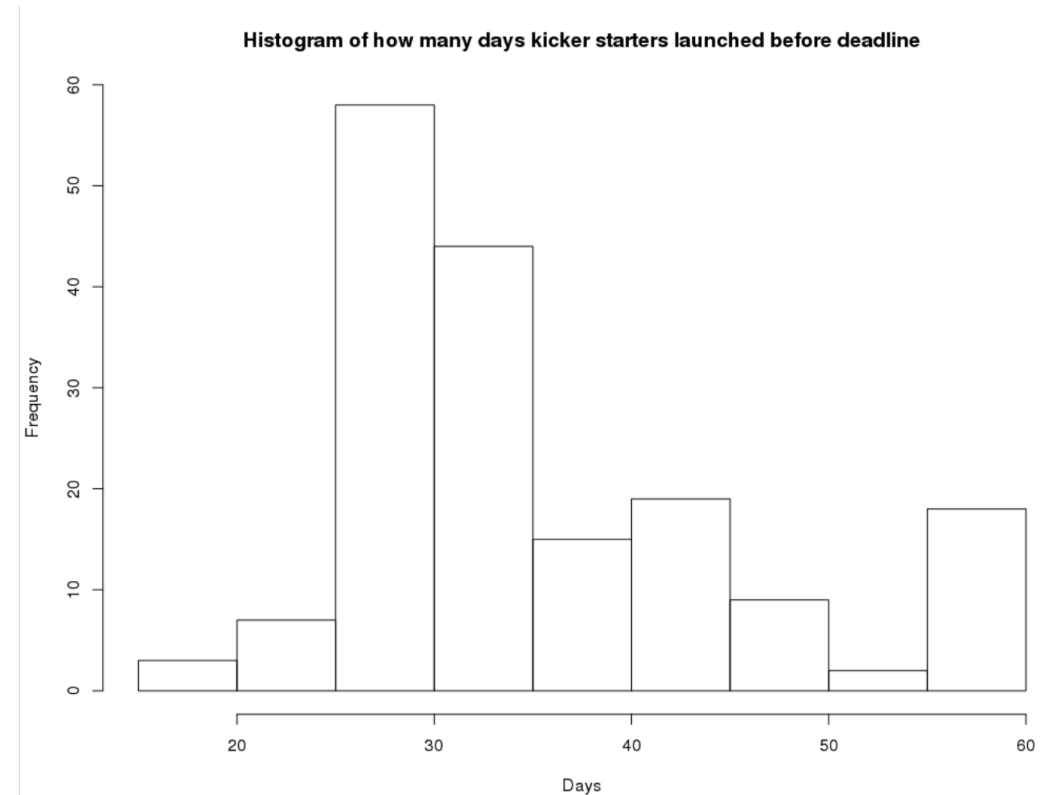
- Every Kickstarter launched before their target deadline

Implementation of Features Not Mentioned in Specification

Regular



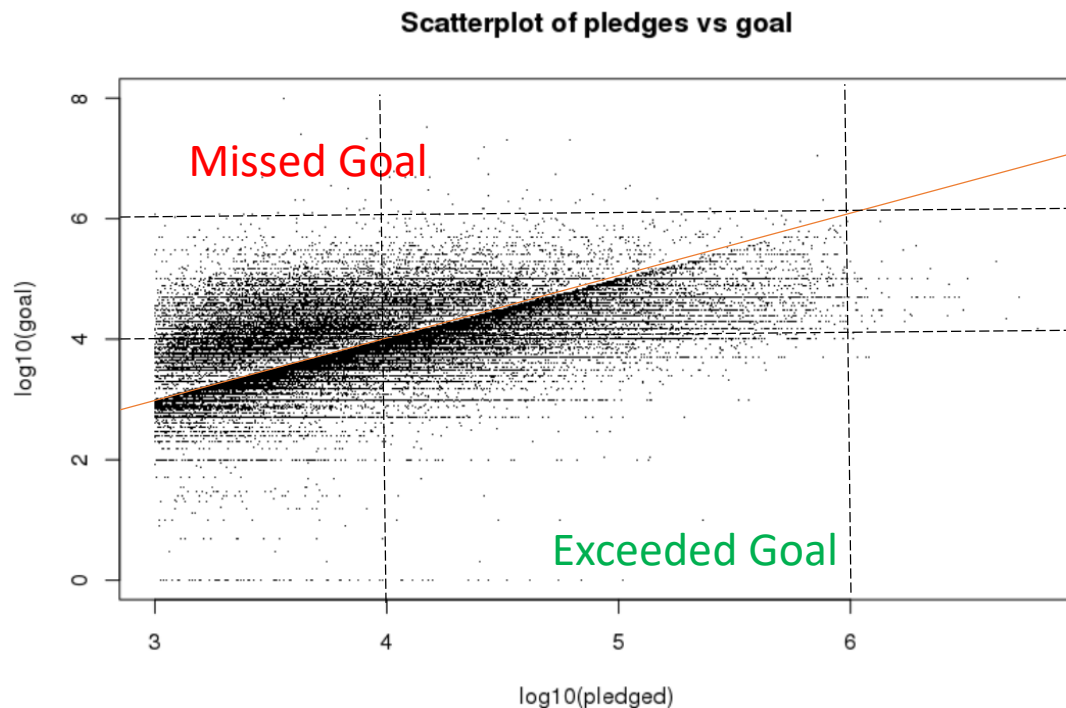
Unicorn



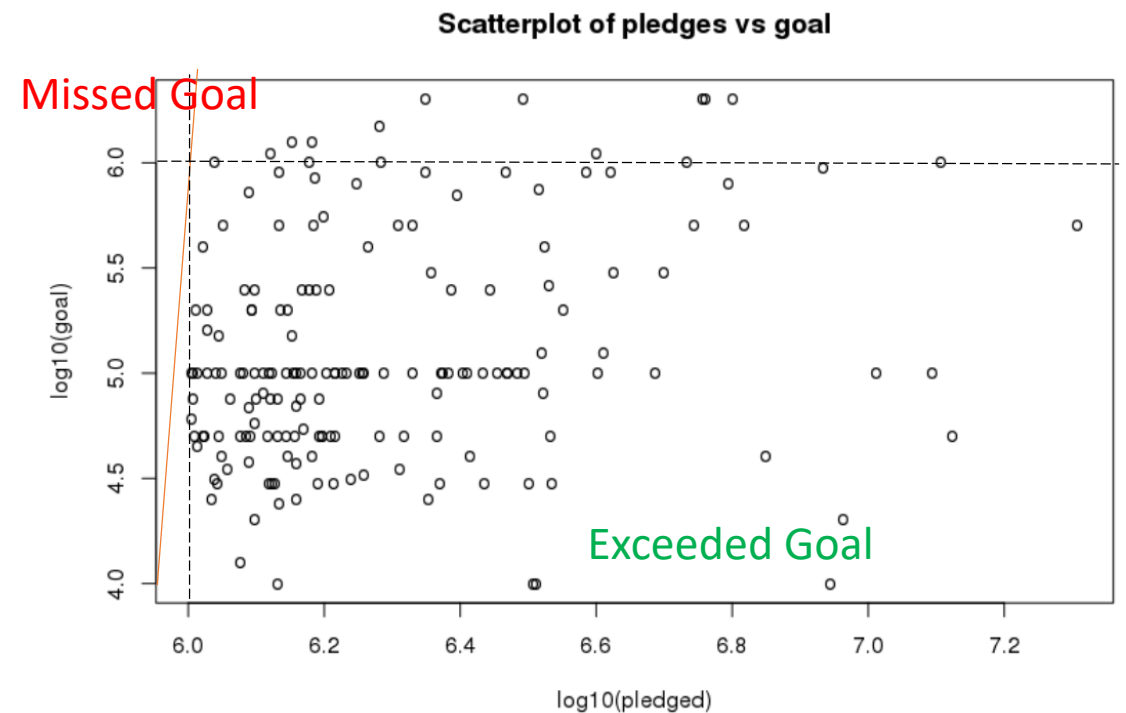
- Regular KickStarters launched approx. 30 days prior to deadline
- Unicorn KickStarters launched approx 35-40 days prior to deadline

Implementation of Features Not Mentioned in Specification

Regular



Unicorn



- A higher percentage of Regular KickStarters met or exceeded their target Pledge Goal
- Every Unicorn KickStarter exceeded their target Pledge Goal

Thank You