



# THE DYNAMICS OF REWARD- BASED CROWDFUNDING

— predicting Kickstarter campaign success —

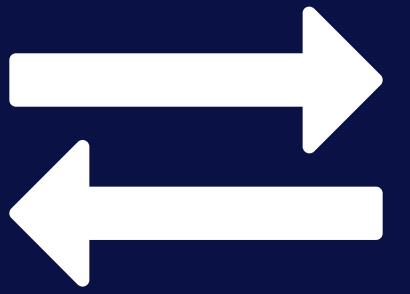
R.J. Klaasse Bos



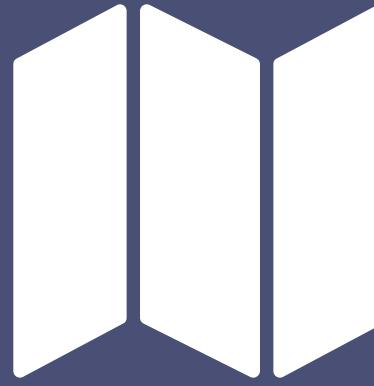
# Table of Contents

---

1. Hypotheses



2. Method



3. Results



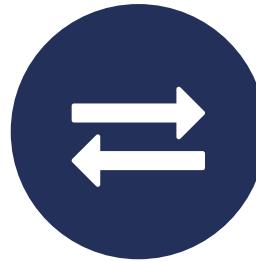
# What Is **KICKSTARTER** About?



**CREATOR**  
(INITIATOR)

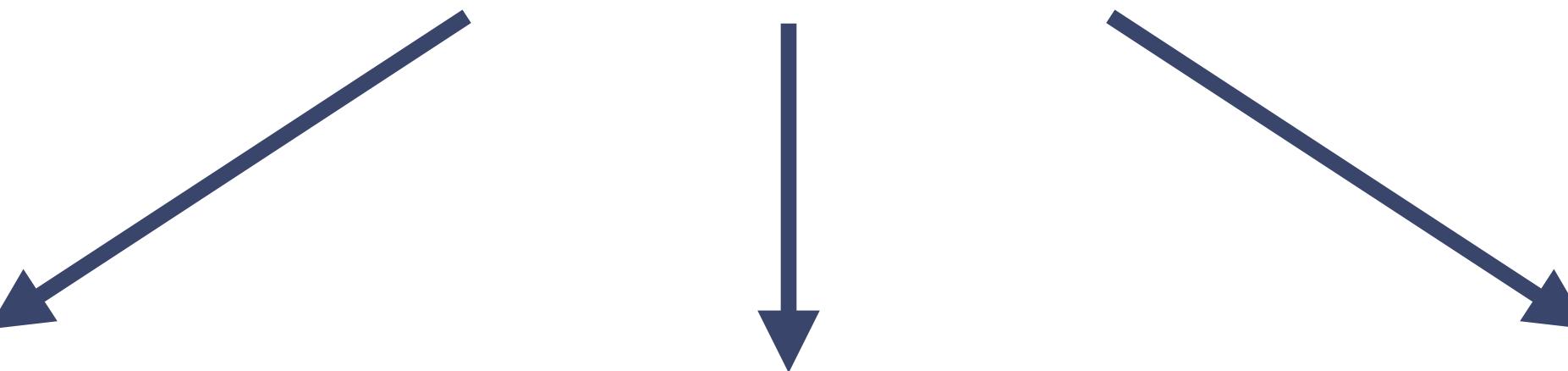


**BACKERS**  
(CROWD)

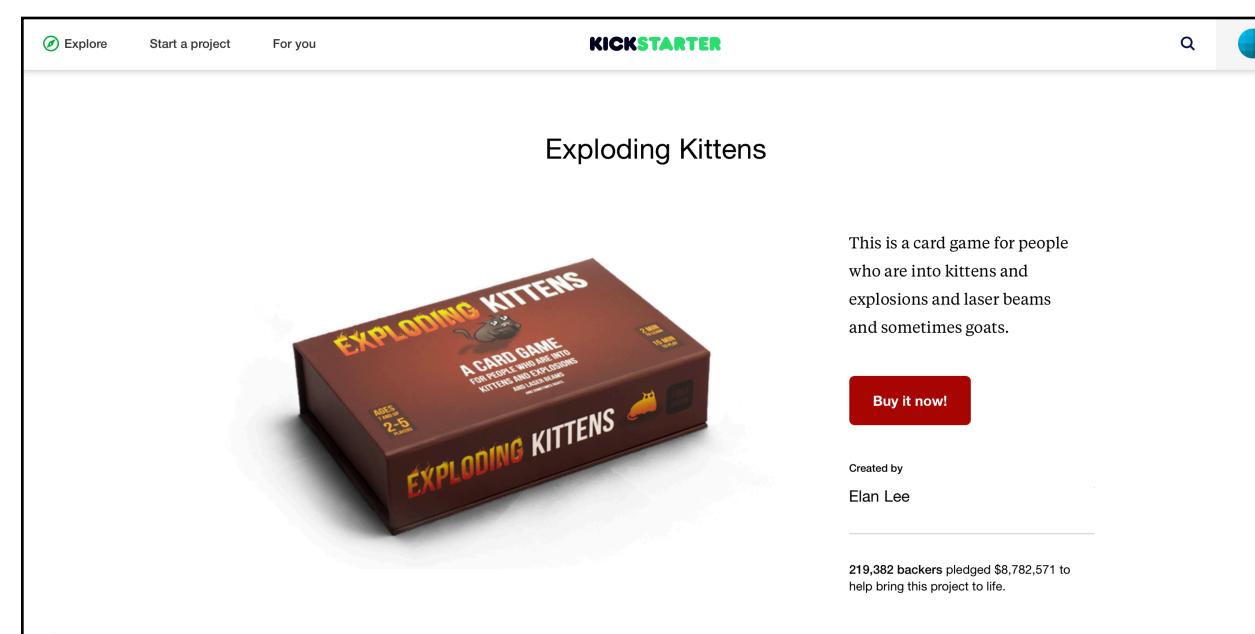


# Literature Gap

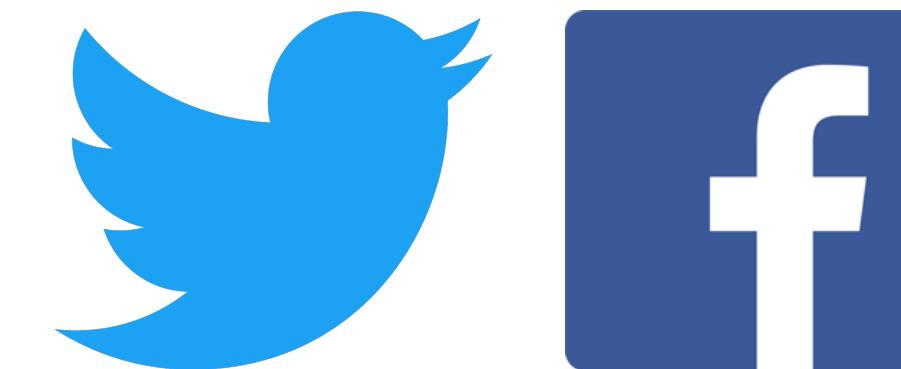
## Prior Research



### 1. Kickstarter Features

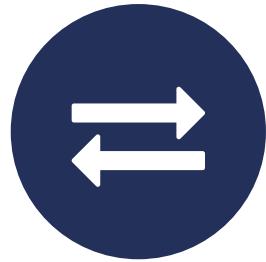


### 2. Social Features



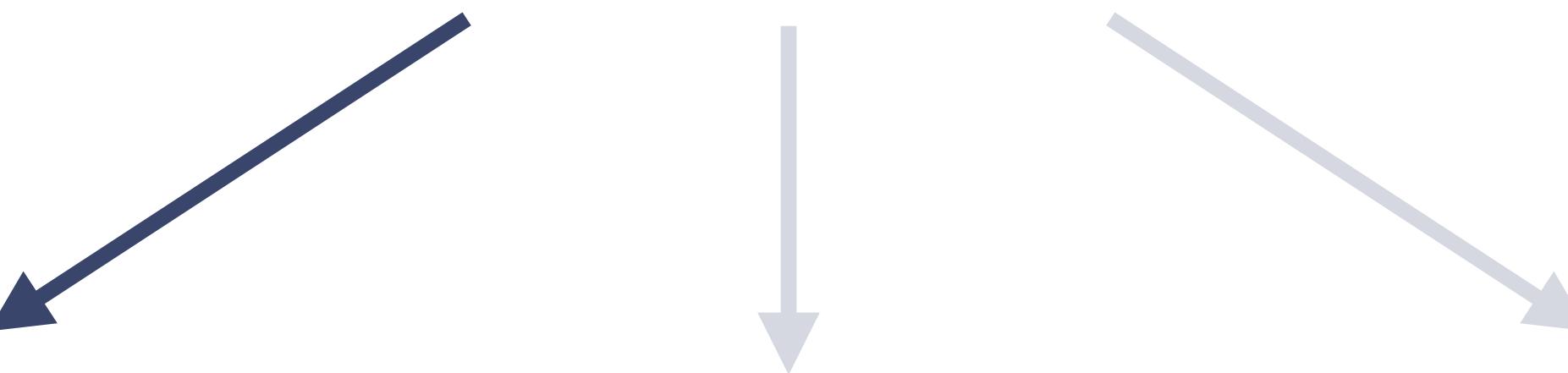
### 3. Linguistic Features

**Pebble** is the first watch built for the 21st century. It's infinitely customizable, with beautiful downloadable watchfaces and useful internet-connected apps. Pebble connects to iPhone and Android smartphones using Bluetooth, alerting you with a silent vibration to incoming calls, emails and messages. While designing Pebble, we strove to create a minimalist yet fashionable product that seamlessly blends into everyday life.

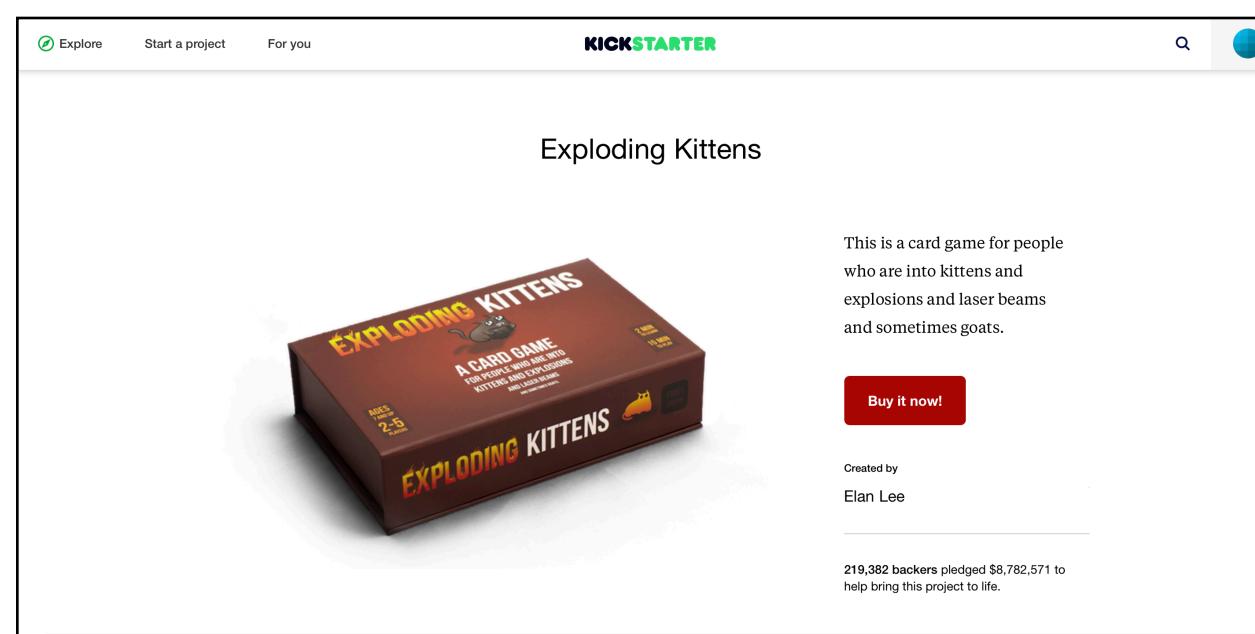


# Literature Gap

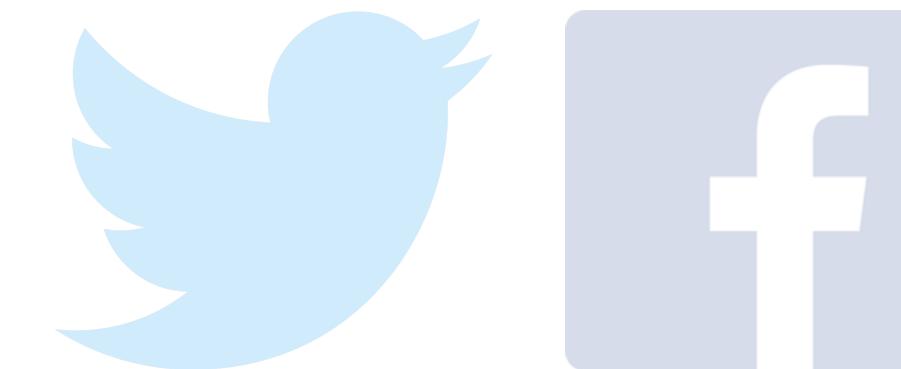
## Prior Research



### 1. Kickstarter Features



### 2. Social Features



### 3. Linguistic Features

**Pebble** is the first watch built for the 21st century. It's infinitely customizable, with beautiful downloadable watchfaces and useful internet-connected apps. Pebble connects to iPhone and Android smartphones using Bluetooth, alerting you with a silent vibration to incoming calls, emails and messages. While designing Pebble, we strove to create a minimalist yet fashionable product that seamlessly blends into everyday life.

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

Campaign

FAQ <sup>19</sup>

Updates <sup>33</sup>

Comments <sup>110,620</sup>

Community



[Explore](#)[Start a project](#)[For you](#)**KICKSTARTER**[Search](#)

# Exploding Kittens



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

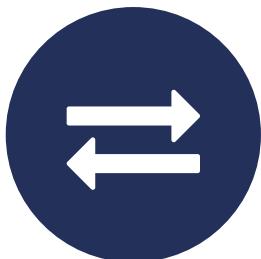
[Buy it now!](#)

Created by

Elan Lee



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



# Goals

---

## Attribute

Presence of video

Location creator

(Sub)category

Goal

#Updates

#Comments

#Images

#Words in project description

#Pledge tiers

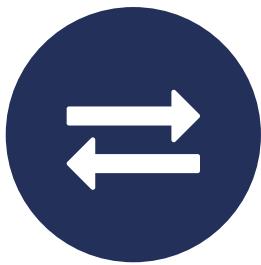
Duration

#Facebook Friends

#Projects Created

#Projects Backed

1. Discover **unexamined attributes** which contribute to campaign success
2. Build a classifier for campaign success



# Literature Gap

## Attribute

Presence of video

Location creator

(Sub)category

Goal

#Updates

#Comments

#Images

#Words in project description

#Pledge tiers

Duration

#Facebook Friends

#Projects Created

#Projects Backed

## Estimated Delivery (H1)

Pledge \$20 or more

THE EXPLODING KITTENS DECK

One copy of Exploding Kittens. (Ages 7+)

-To order multiple decks, just add \$20 for each extra deck you'd like. (extra shipping charges may apply)

ESTIMATED DELIVERY

Jul 2015

SHIPS TO

Anywhere in the world

15,505 backers

Estimated  
Delivery (H1)



Success Rate  
Kickstarter Project

Fast delivery affects customer **loyalty** (Dotcom, 2016)

63% of all shoppers find it important to receive their order in  
the **shortest amount of time** possible (Dotcom, 2016)

**Shipping speed** is one of the top 5 factors consumers take into  
account when comparison shopping (comScore, 2012)

# Mean Estimated Delivery (H1)



# Success Rate Kickstarter Project

## Pledge \$20 or more

### THE EXPLODING KITTENS DECK

One copy of Exploding Kittens. (Ages 7+)

-To order multiple decks, just add \$20 for each extra deck you'd like. (extra shipping charges may apply)

ESTIMATED DELIVERY  
Jul 2015

SHIPS TO  
Anywhere in the world

15,505 backers

## Pledge \$35 or more

### THE NSFW DECK

One copy of the Exploding Kittens deck PLUS one copy of the NSFW deck. This is a deck of bonus cards that were too horrible/incredible to include in the kid friendly version. (Ages 30+)

-To order multiple pairs of decks, just add \$35 for each extra pair of decks you'd like. (extra shipping charges may apply)

ESTIMATED DELIVERY  
Jul 2015

SHIPS TO  
Anywhere in the world

202,934 backers

## Pledge \$100 or more

### THE COLLECTOR'S DECK

Two copies of both decks from the previous reward but one of each will be signed by the creators of the game. (So you don't have to ruin the fancy signed decks with your filthy game playing hands.)

ESTIMATED DELIVERY  
Jul 2015

SHIPS TO  
Anywhere in the world

Reward no longer available

200 backers

## Pledge \$500 or more

### THE DECK OF LEGENDS

Everything from the previous reward PLUS each of the game creators will draw you a custom card. (We apologize in advance for Elan's drawing skills.)

ESTIMATED DELIVERY  
Aug 2015

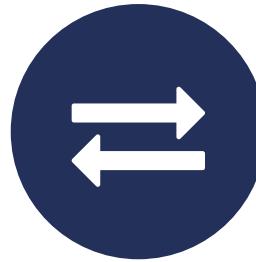
SHIPS TO  
Anywhere in the world

Reward no longer available

5 backers

## Funding period

Jan 20 2015 - Feb 20 2015 (30 days)



# Literature Gap

---

## Attribute

Presence of video

Location creator

(Sub)category

Goal

#Updates

#Comments

#Images

#Words in project description

#Pledge tiers

Duration

#Facebook Friends

#Projects Created

#Projects Backed

## Past Success Rate (H2)

#Previous Successful campaigns

---

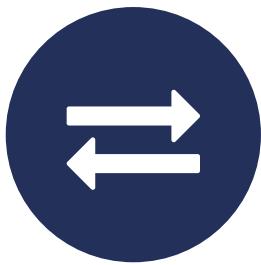
#Previous campaigns

Past Success  
Rate (H2)



Success Rate  
Kickstarter Project

Prior crowdfunding experience is likely related to the creator's trustworthiness (Chen et al.)



# Literature Gap

## Attribute

Presence of video

Location creator

(Sub)category

Goal

#Updates

#Comments

#Images

#Words in project description

#Pledge tiers

Duration

#Facebook Friends

#Projects Created

#Projects Backed

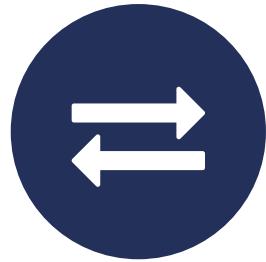
## Adoption Speed (H3)



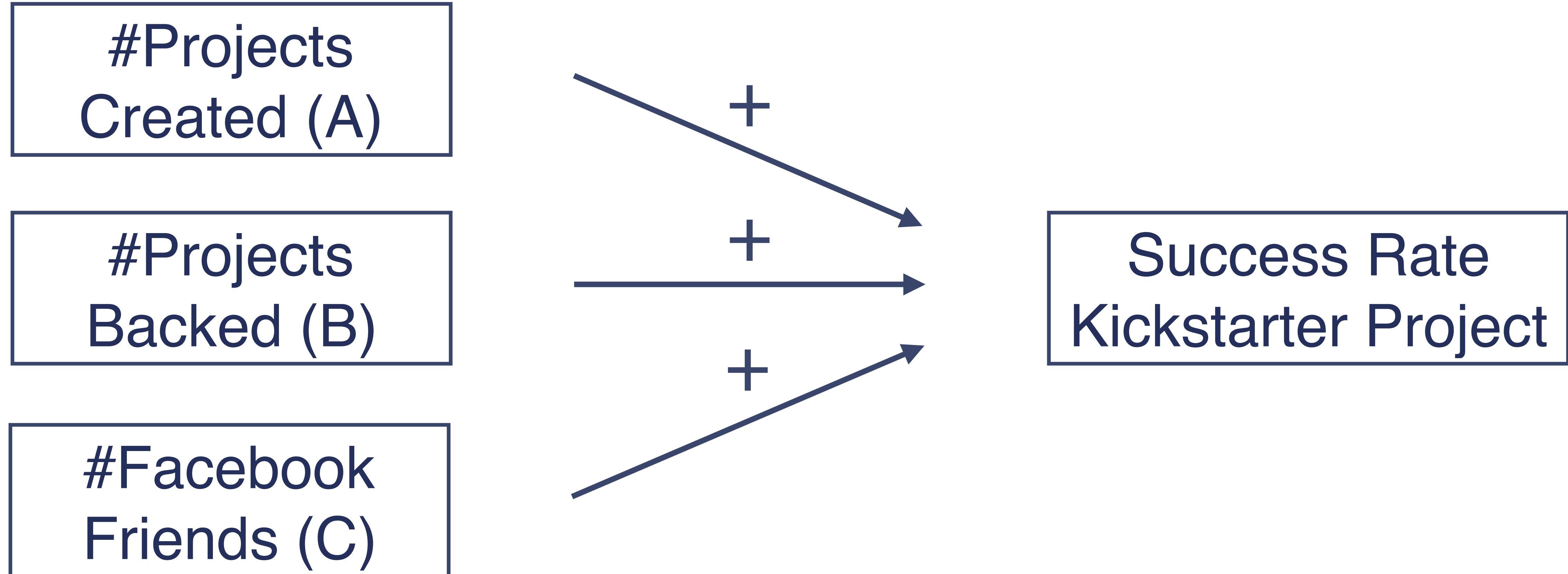
**Elan Lee**

Backed 306 projects · Joined Jun 2010

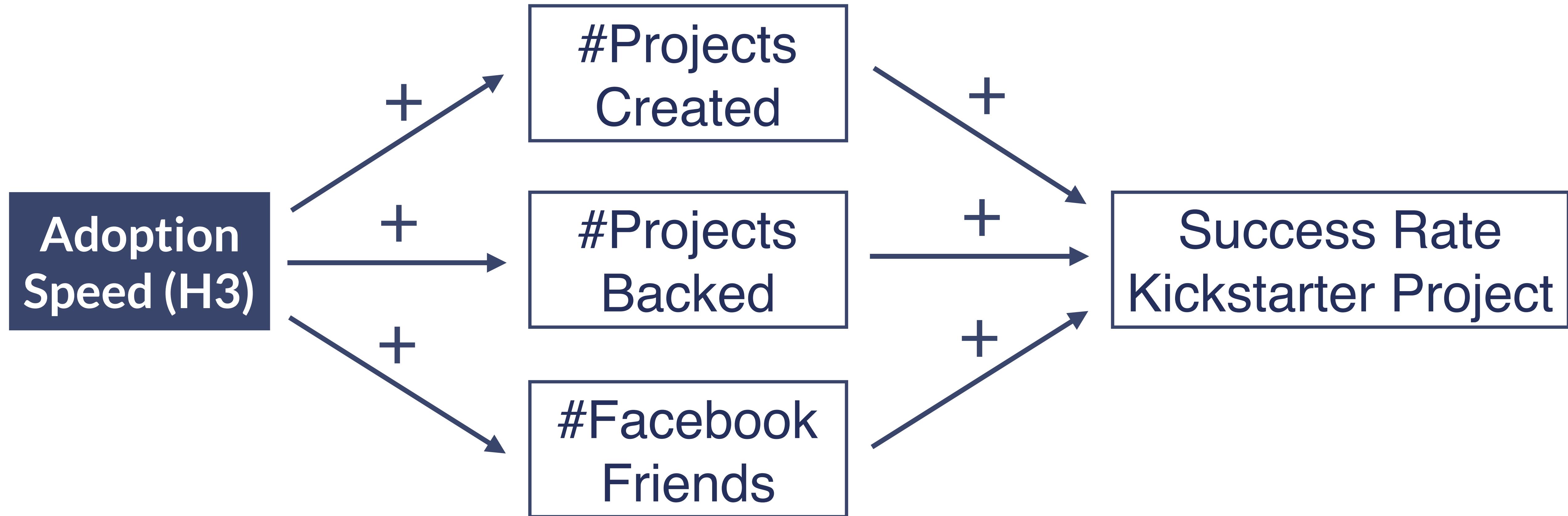
Follow

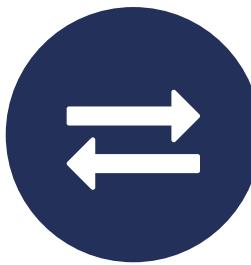


# Prior Research

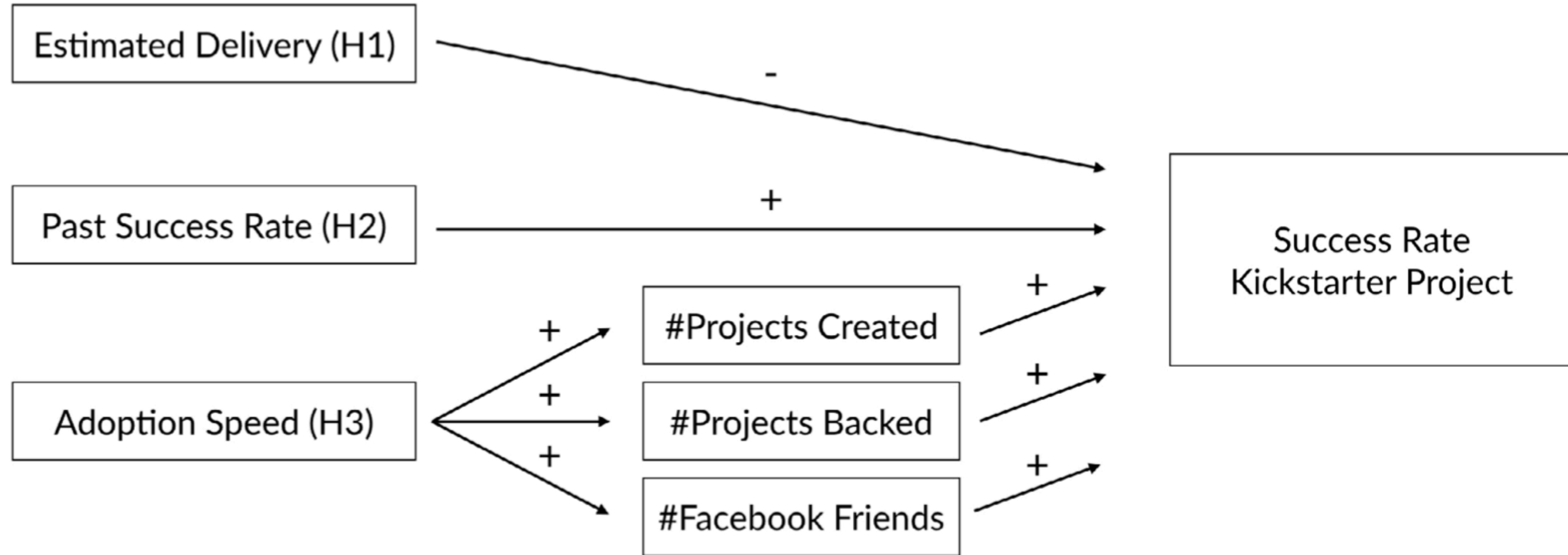


A: Chen et al. (n.d.); B: Chen et al. (n.d.); C: Mollick (2014)





# Conceptual Model





# Methodology | Scrape Data

n= 3726

The screenshot shows the Kickstarter main page for the game "Exploding Kittens". It features a large image of the game box, which is brown with yellow and white text. The title "EXPLDING KITTENS" is prominently displayed. Below the box, there is a brief description: "This is a card game for people who are into kittens and explosions and laser beams and sometimes goats." A red "Buy it now!" button is visible. At the bottom, it says "219,382 backers pledged \$8,782,571 to help bring this project to life."

MAIN PAGE

The screenshot shows the bio page for Elan Lee. It includes a green checkmark icon next to her name, indicating she is a verified user. Below that, it says "Last login Jun 23 2017". There is a Facebook icon followed by "1,927 friends". Underneath, it shows "2 created · 306 backed". A blue "Contact me" button is at the bottom.

BIO PAGE

The screenshot shows Elan Lee's profile page. It features a circular profile picture of a man sitting on a stool. Below the picture, her name "Elan Lee" is displayed in bold black text. To the right, it says "Backed 306 projects · Joined Jun 2010". There are links for "About", "Backed 306", "Created 2", and "Comments 134". At the bottom, there is a thumbnail for another project called "EARS VS DRTES" with the description "Bears vs Babies - A Card Game".

PROFILE PAGE



# Methodology | Data Preprocessing (1)

Attribute	Before	After
City	Hollywood, Los Angeles, CA	Los Angeles
State	Hollywood, Los Angeles, CA	California
Number of Backers	2 backers	2
Goal	AU\$ 1,000.00	\$757.50
Main Video	"You'll need an HTML5 capable browser to see this content."	1
....	...	...
+ Derived features		
+ Remove incomplete cases		



# Methodology | Data Preprocessing (2)

The screenshot shows a web browser window with the URL `royklnl84.eightyfour.axc.nl/kickstarter.html`. The page content is as follows:

## Number of backers

This step involves removing character data from a string. Or in concrete terms: from "2 backers" to 2 . Singular (i.e. "backer" ) and plural forms (i.e. "backers" ) should be taken into account.

```
generalStats$totalNumberBackers = gsub( ",", "", generalStats$totalNumberBackers)
generalStats$totalNumberBackers = gsub( " backer", "", generalStats$totalNumberBackers)
generalStats$totalNumberBackers = gsub( "s", "", generalStats$totalNumberBackers)
```

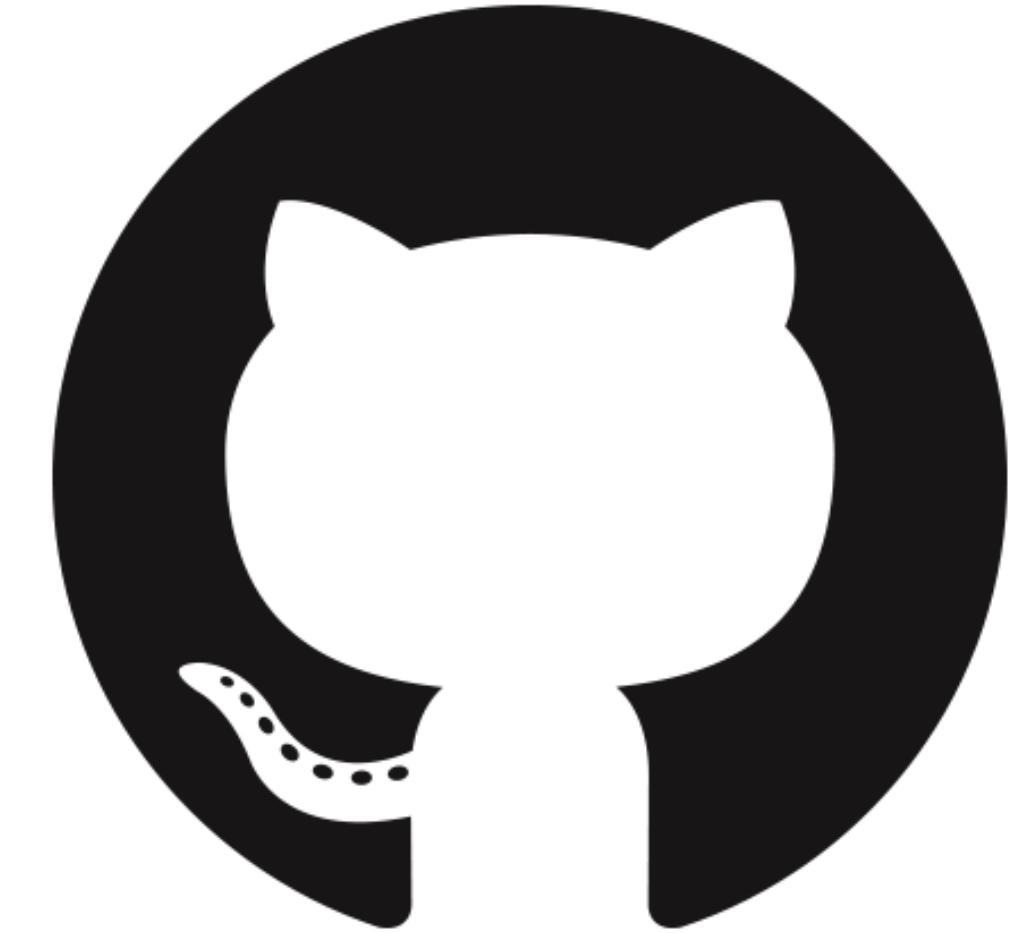
For the perks the same process is repeated. This time though, additional perk parameters should be taken into account (i.e. "Limited" and "Reward no longer available").

```
perkStats$numBackers = gsub( ",", "", perkStats$numBackers)
perkStats$numBackers = gsub( "Limited ", "", perkStats$numBackers)
perkStats$numBackers = gsub( "Reward no longer available ", "", perkStats$numBackers)
perkStats$numBackers = gsub( " backer", "", perkStats$numBackers)
perkStats$numBackers = gsub( "s", "", perkStats$numBackers)
```

## Converting currencies

As stated before, all columns containing monetary data should be converted to a standard unit or baseline. That is why non-USA currencies are multiplied by their exchange rates relative to USD (\$). Note, the exchanges rates below were determined on the 13th of April 2017 and will be assumed to remain unchanged throughout the entire campaign duration (even though that is obviously not the case).

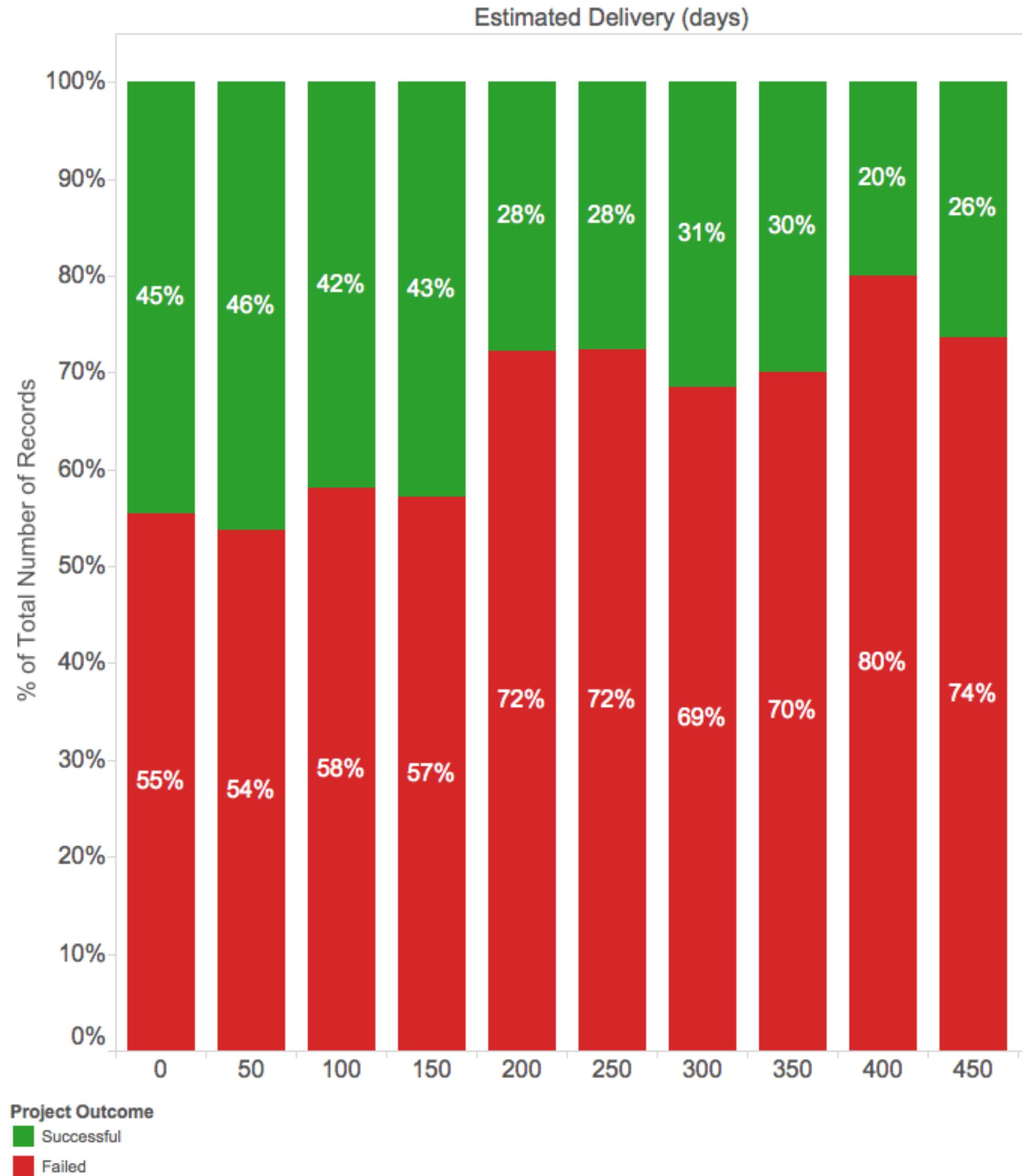
```
aud_exch = 0.7575 #Australian Dollar
cad_exch = 0.7511 #Canadian Dollar
```



For details see: [bit.ly/KickstarterPreprocessing](https://bit.ly/KickstarterPreprocessing) + link to cleaned data



# Estimated Delivery (H1)





# Estimated Delivery (H1)

Model 1 - (Only including "Estimated Delivery(mean)")			
Attribute	Estimate	Std. Error	P-value
Estimated Delivery (mean)	-2.23e-03	5.12e-04	1.31e-05***
Goal	-6.41e-05	4.74e-06	<2e-16***
Duration	-2.30e-02	4.29e-03	8.14e-08***
#Updates	2.85e-01	1.85e-02	<2e-16***
#Comments	6.20e-02	6.88e-03	<2e-16***
#Images	-2.20e-03	4.45e-03	0.62
#Perks	9.94e-02	1.05e-02	<2e-16***

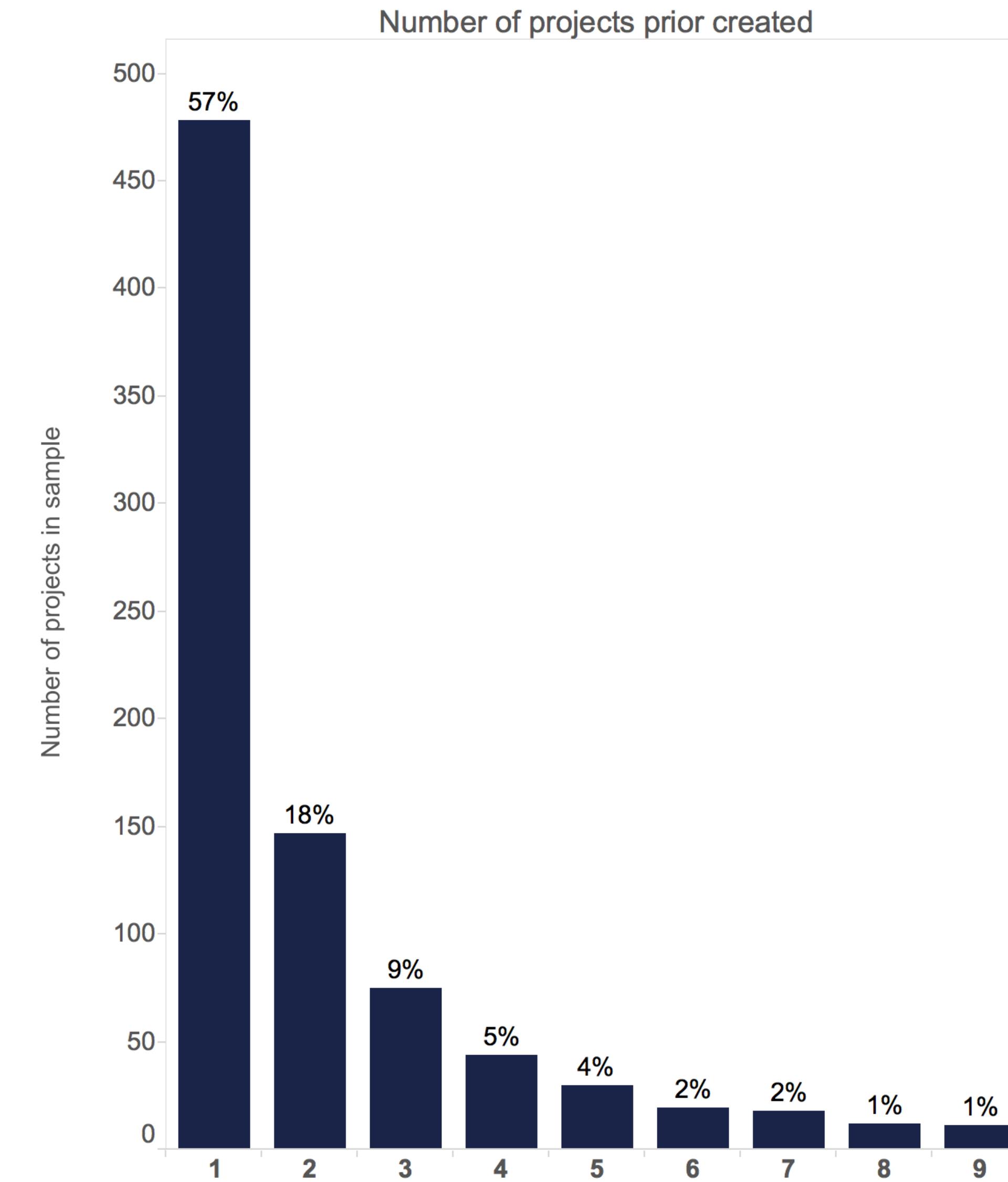
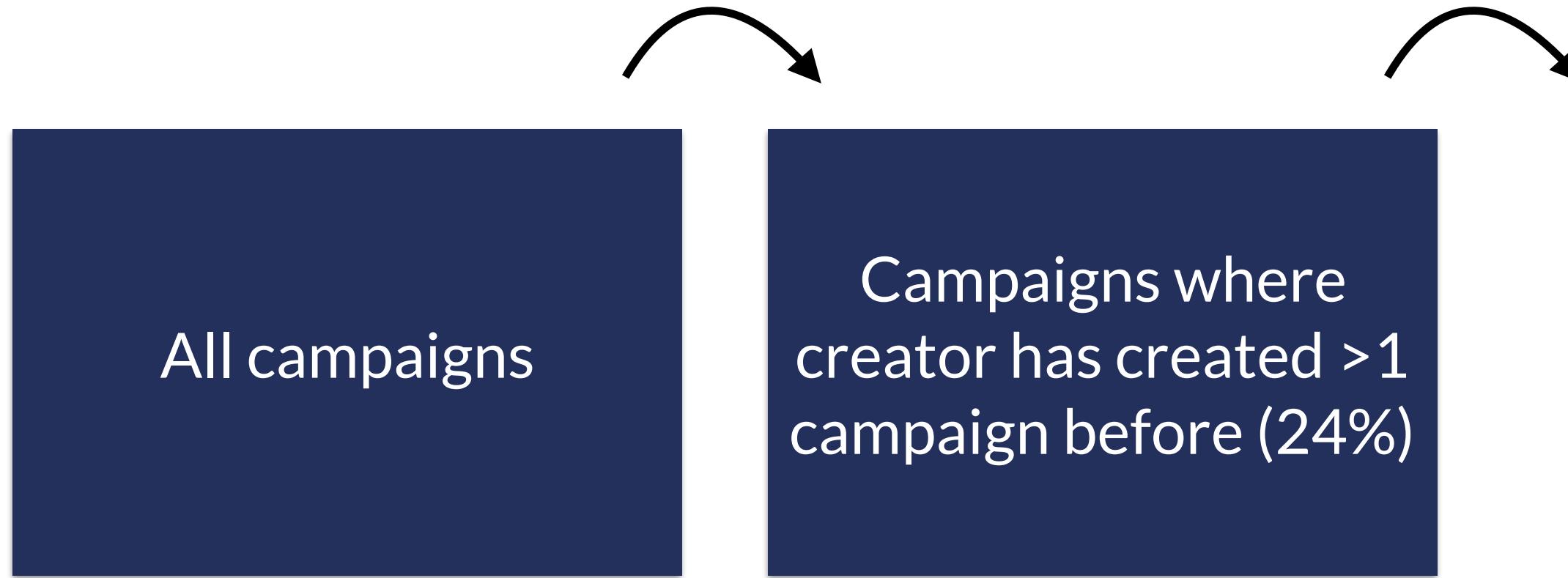
Dependent variable: *project success*

Model 2 - (Only including "Estimated Delivery(individual)")			
Attribute	Estimate	Std. Error	P-value
Estimated Delivery (individual)	-1.21e-04	1.54e-01	0.65
Goal	-5.00e-05	3.17e-06	<2e-16***
Duration	1.89e-02	4.08e-03	3.65e-06***
#Updates	2.03e-01	1.35e-02	<2e-16***
#Comments	7.90e-02	7.09e-03	<2e-16***
#Images	2.09e-03	3.65e-03	0.57
#Perks	8.24e-02	7.81e-03	<2e-16***

Dependent variable: *project success*

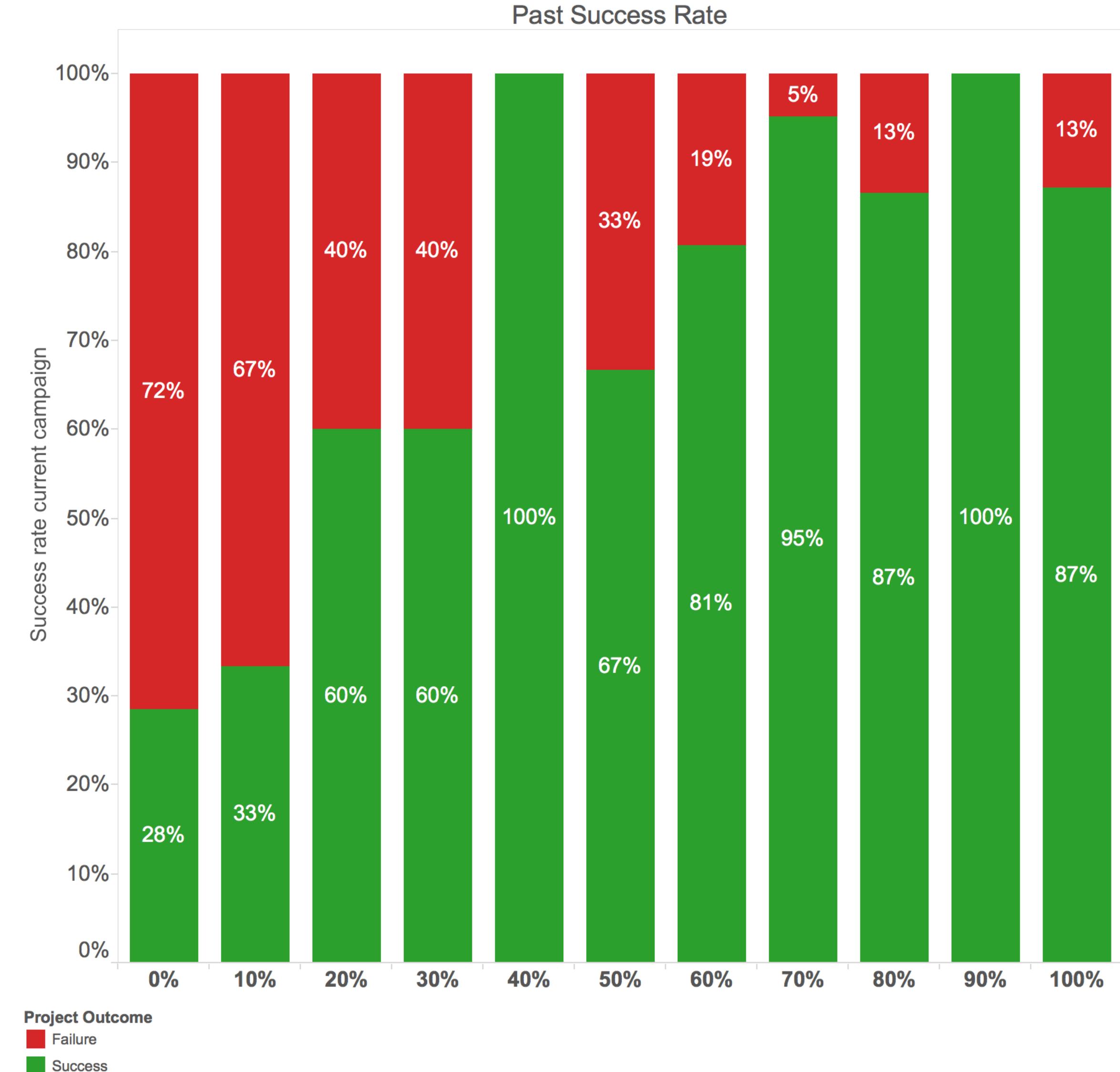
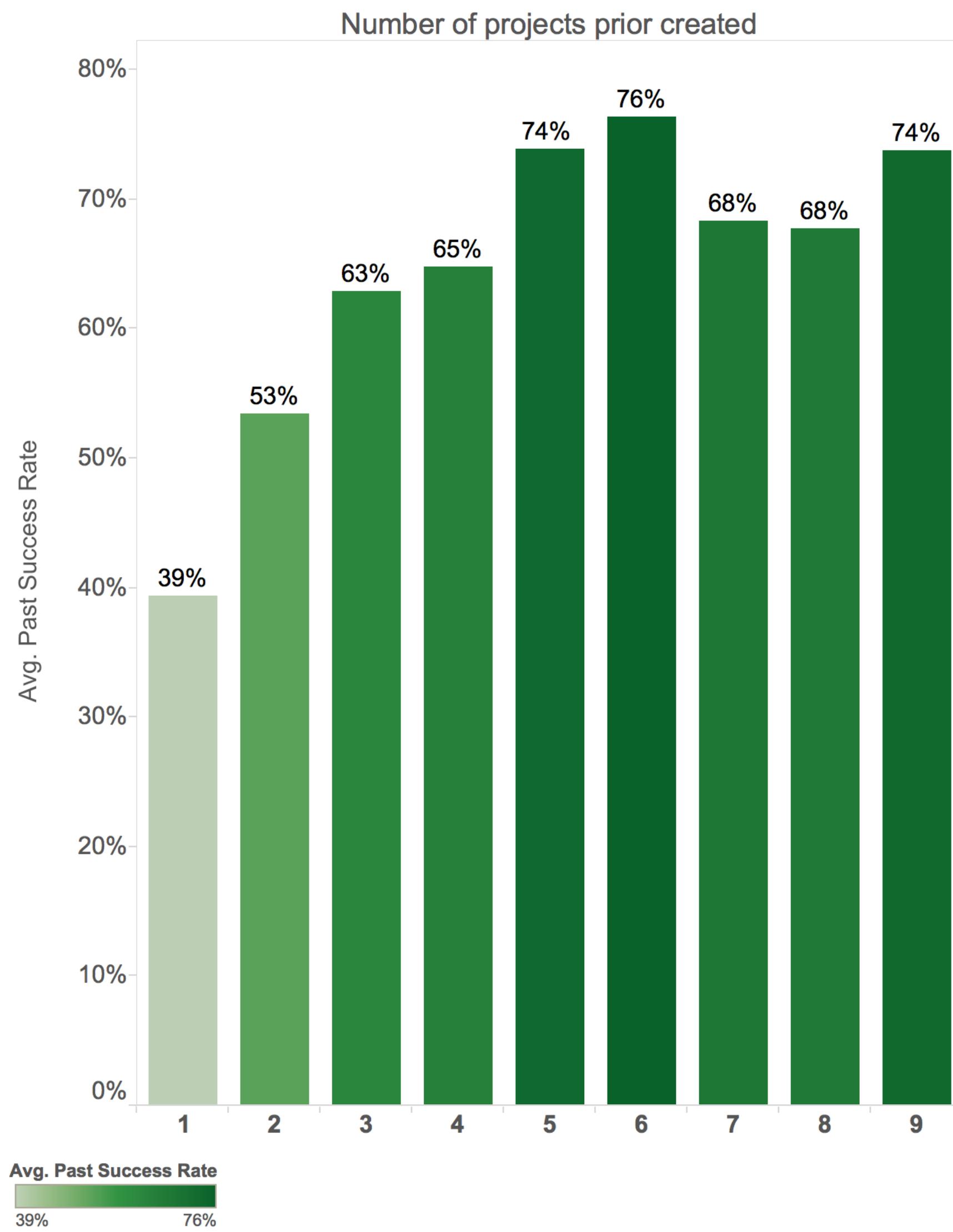


# Past Success Rate (H2)





# Past Success Rate (H2)





# Past Success Rate (H2)

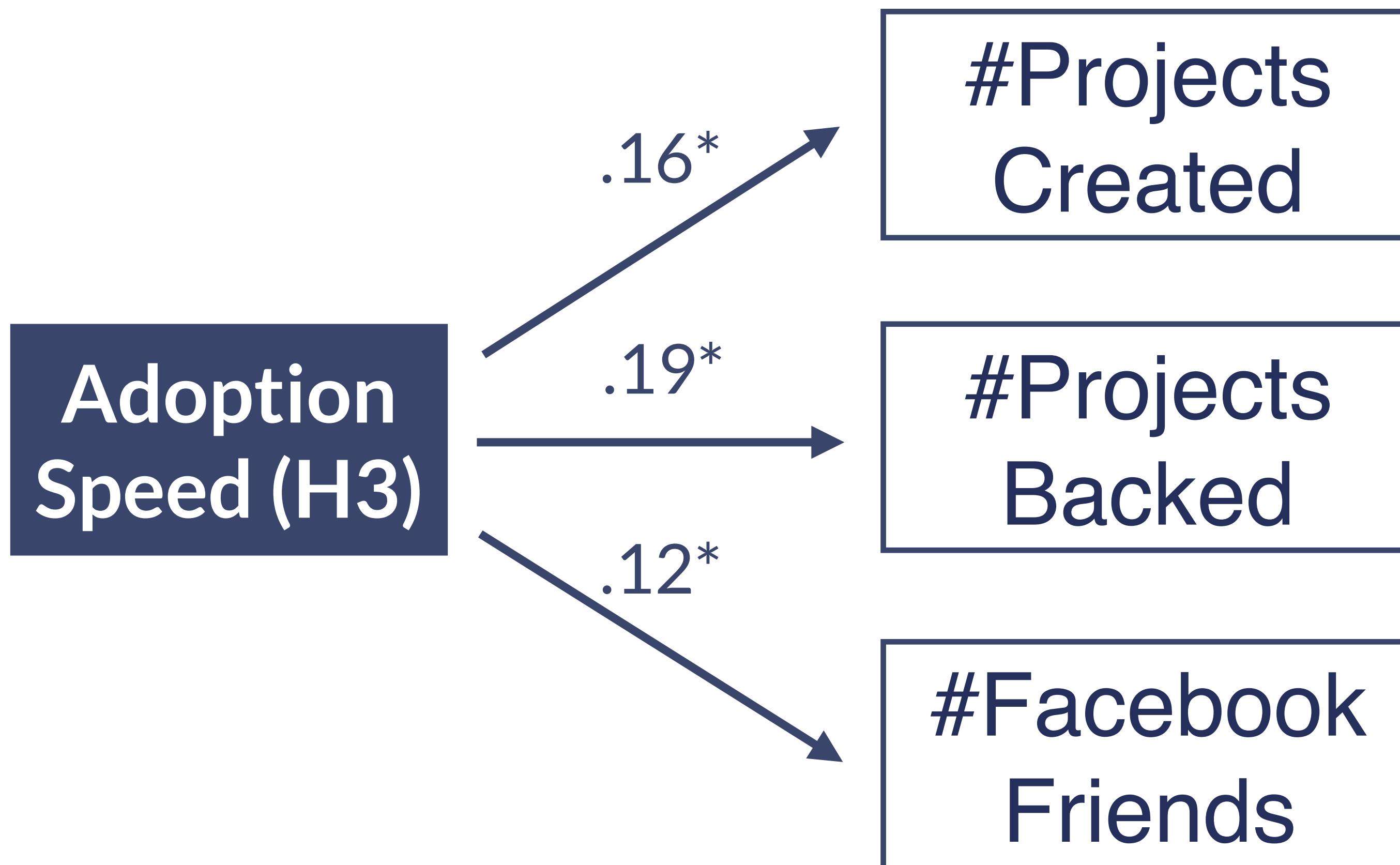
**Model 3** - (Including both “Past Success Rate” and “#Projects Created”)

Attribute	Estimate	Std. Error	P-value
Past Success Rate	2.41e+00	2.30e-01	<2e-16***
#Projects Created	1.31e-01	3.46e-02	1.55e-04***
Goal	-6.46e-05	4.82e-06	<2e-16***
Duration	-2.15e-02	4.43e-03	1.24e-06***
#Updates	2.82e-01	1.88e-02	<2e-16***
#Comments	5.51e-02	6.68e-03	<2e-16***
#Images	1.65e-03	4.55e-03	0.72
#Perks	9.21e-02	1.06e-02	<2e-16***
#Projects Backed	-1.48e-03	2.25e-03	0.51

Dependent variable: *project success*



# Adoption Speed (H3)



Adoption speed vs #Projects Created

Attribute	Estimate	Std. Error	P-value
Adoption speed	0.16	0.020	5.52e-16***
#Projects Backed	0.21	0.019	<2e-16***
#Comments by creator	9.69e-02	1.90e-02	<2e-16***
Facebook connected	0.06	0.036	0.096
#Facebook friends	4.82e-02	1.74e-02	0.0023**
Past Success Rate	0.043	0.018	0.017*

Dependent variable: #Projects Created

Adoption speed vs #Projects Backed

Attribute	Estimate	Std. Error	P-value
Adoption speed	0.19	0.020	<2e-16***
#Projects Created	0.21	0.018	<2e-16***
#Comments by creator	0.10	0.019	3.29e-07***
Facebook connected	-0.037	0.036	0.293
#Facebook friends	-0.0040	0.018	0.819
Past Success Rate	0.10	0.018	1.36e-08***

Dependent variable: #Projects Backed

Adoption speed vs #Facebook Friends

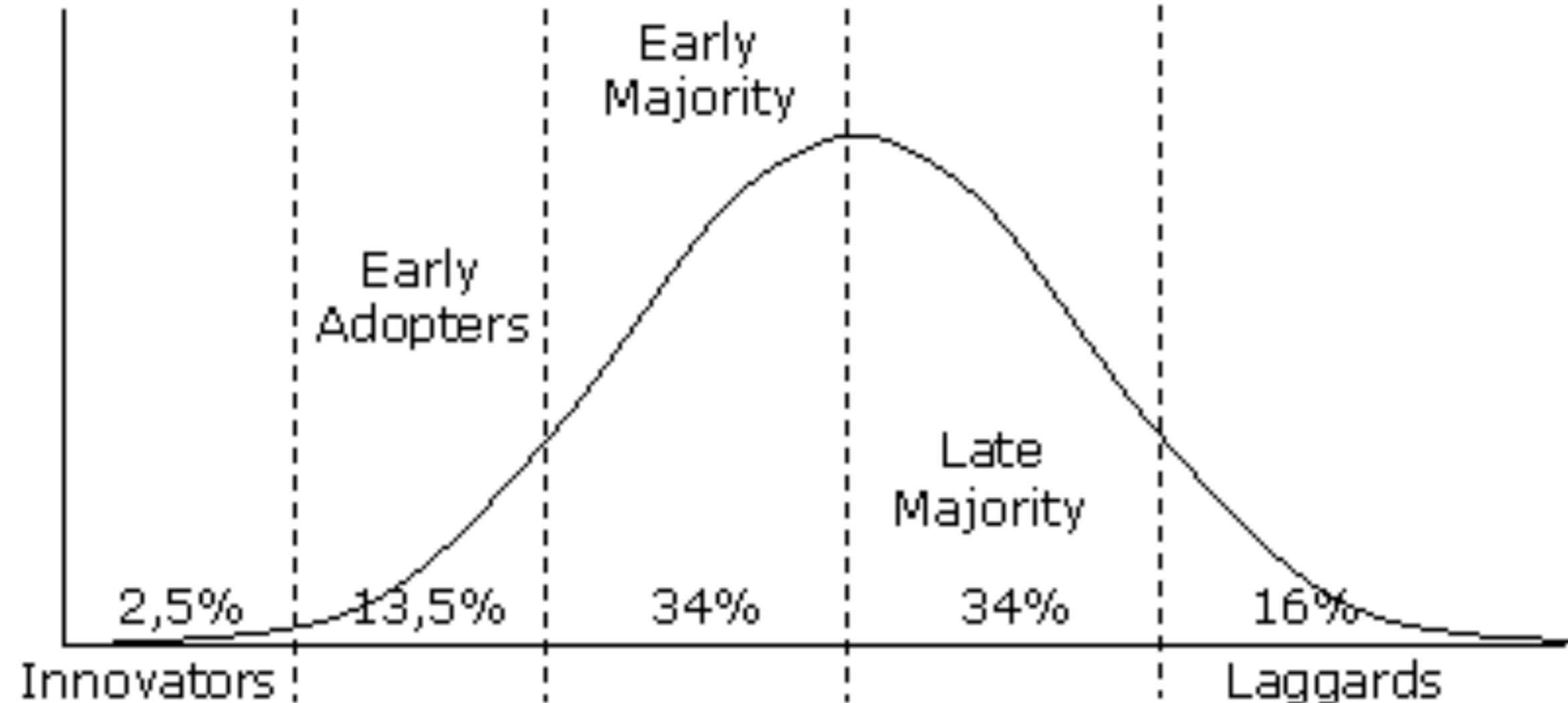
Attribute	Estimate	Std. Error	P-value
Adoption speed	1.19e-01	2.17e-02	5.27e-08***
#Projects Created	5.68e-02	2.06e-02	0.0057**
#Projects Backed	-1.90e-04	2.06e-02	0.99
#Comments by creator	-6.76e-02	2.08e-02	0.0011**
Past Success Rate	4.43e-02	1.95e-02	0.023*

Dependent variable: #Facebook Friends

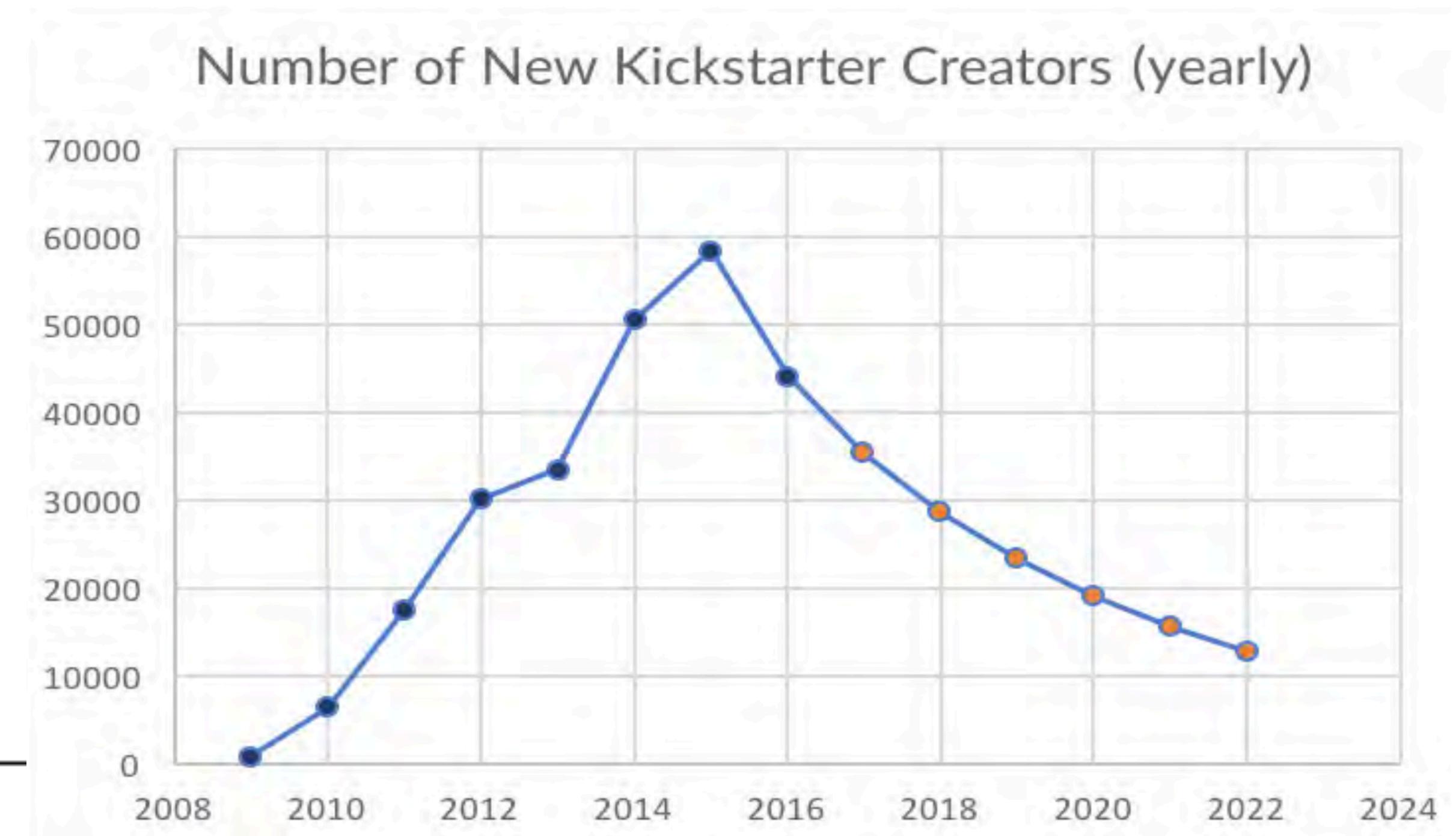


# Adoption Speed (H3)

Rogers Adoption / Innovation Curve

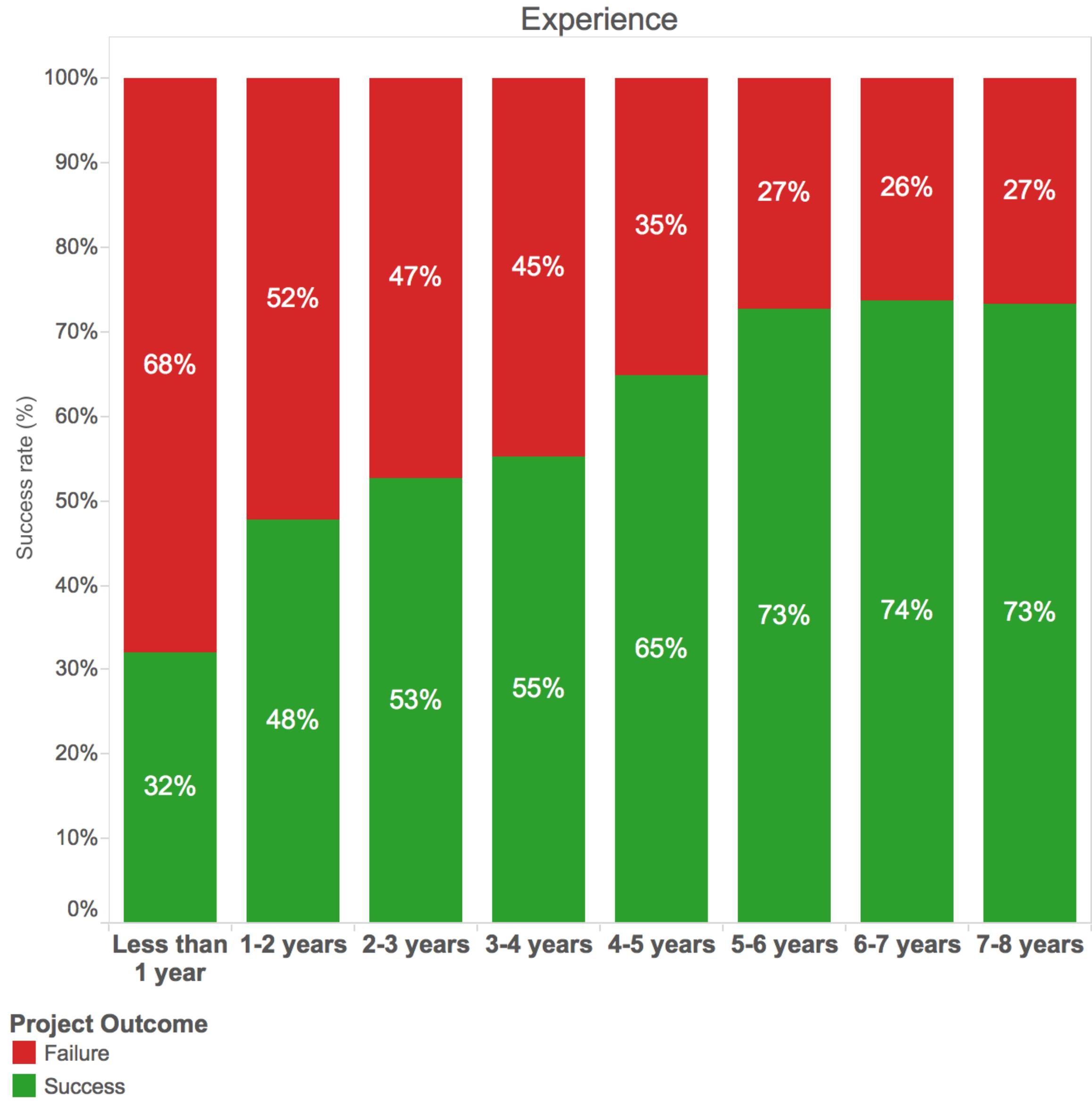


Number of New Kickstarter Creators (yearly)



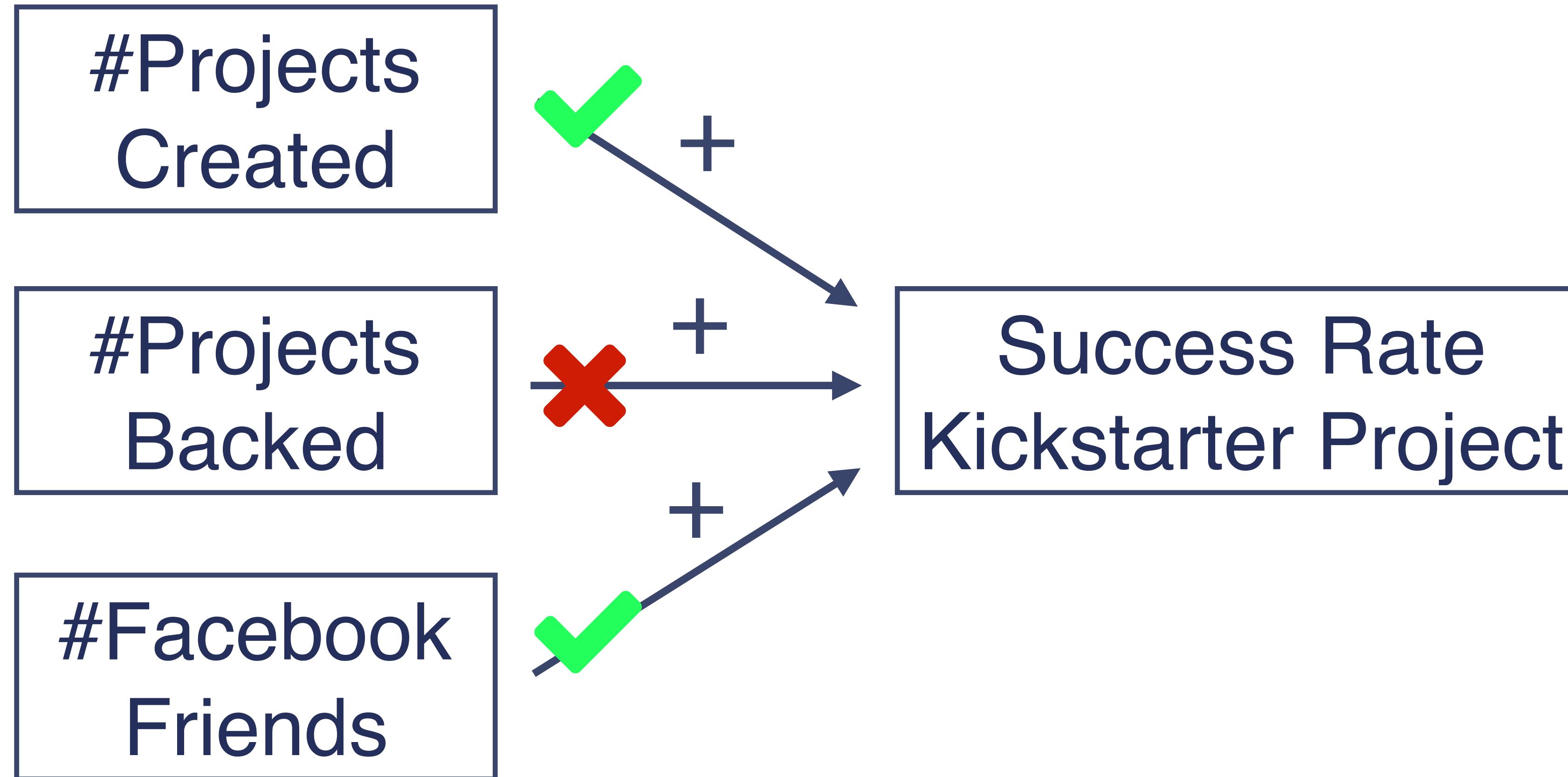


# Adoption Speed (H3)



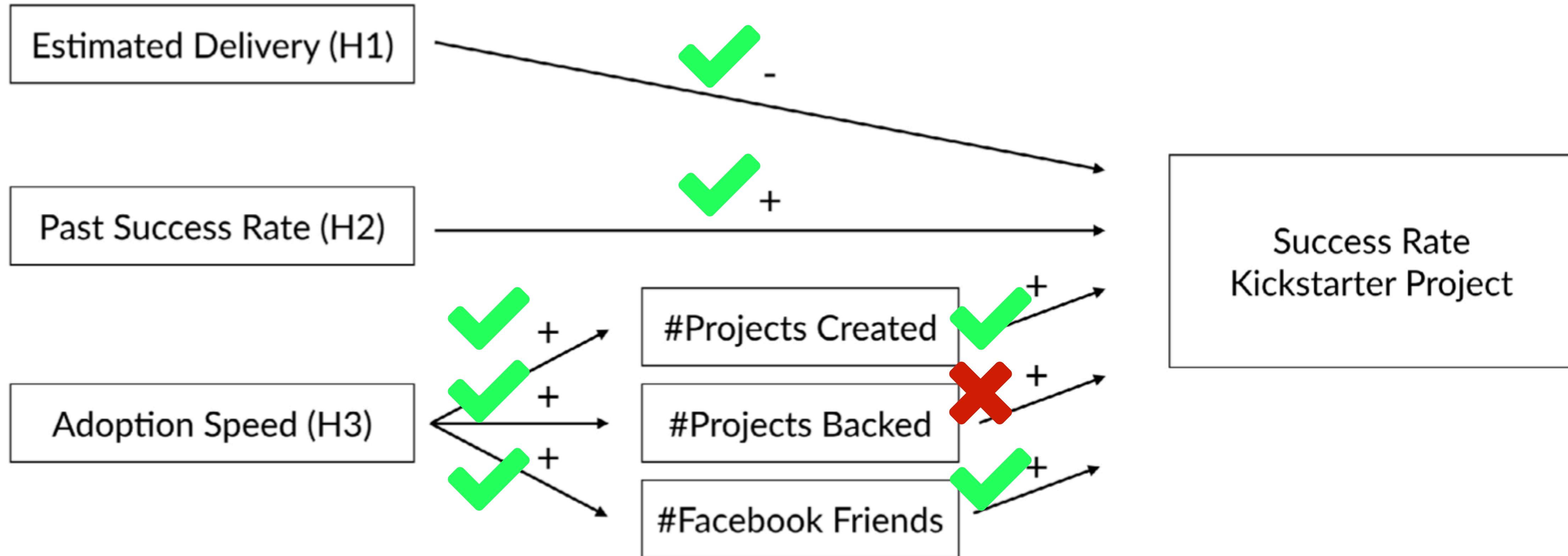


## Adoption Speed (H3)





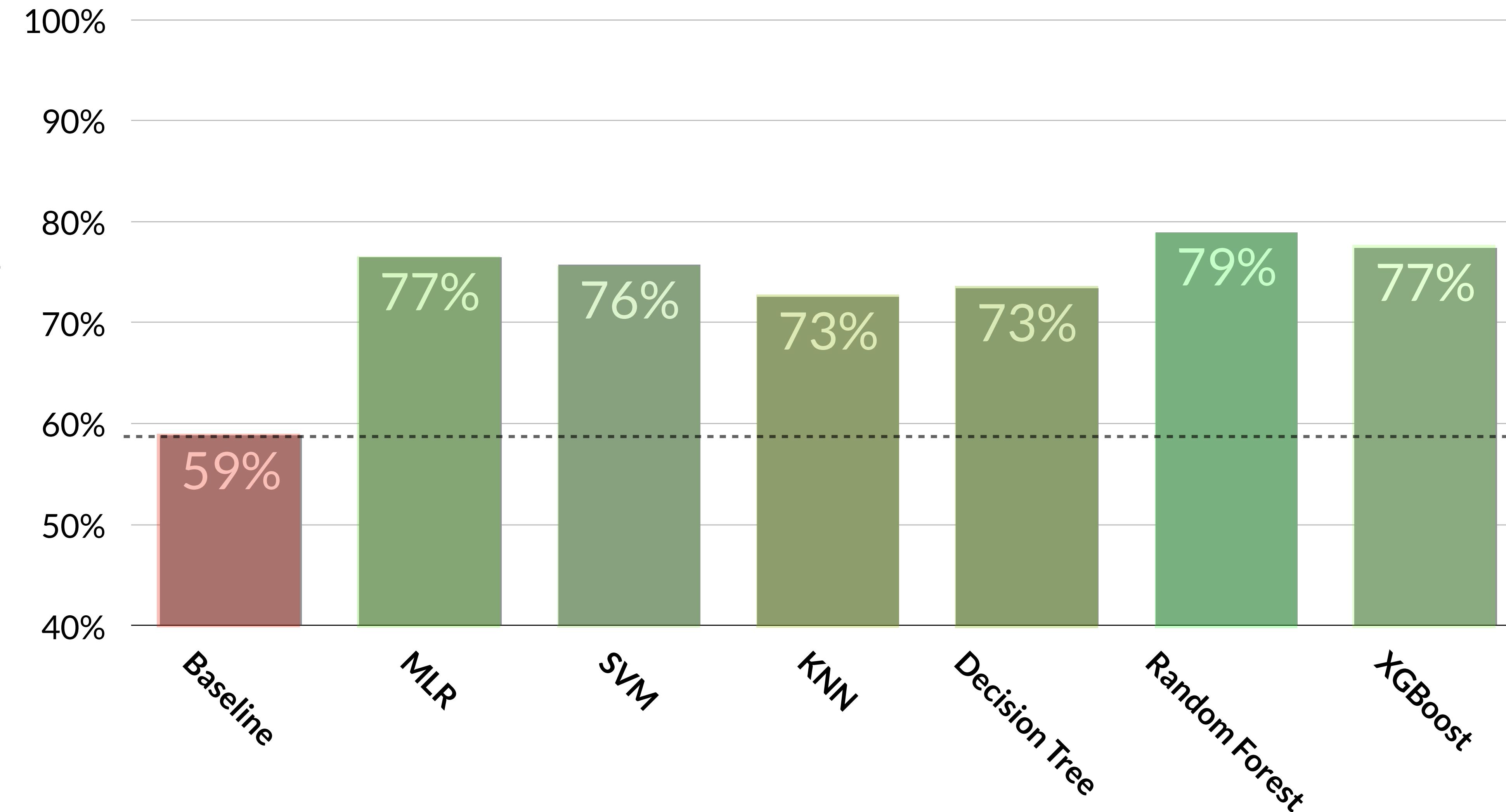
# Results Hypotheses





# Predicting Campaign Success

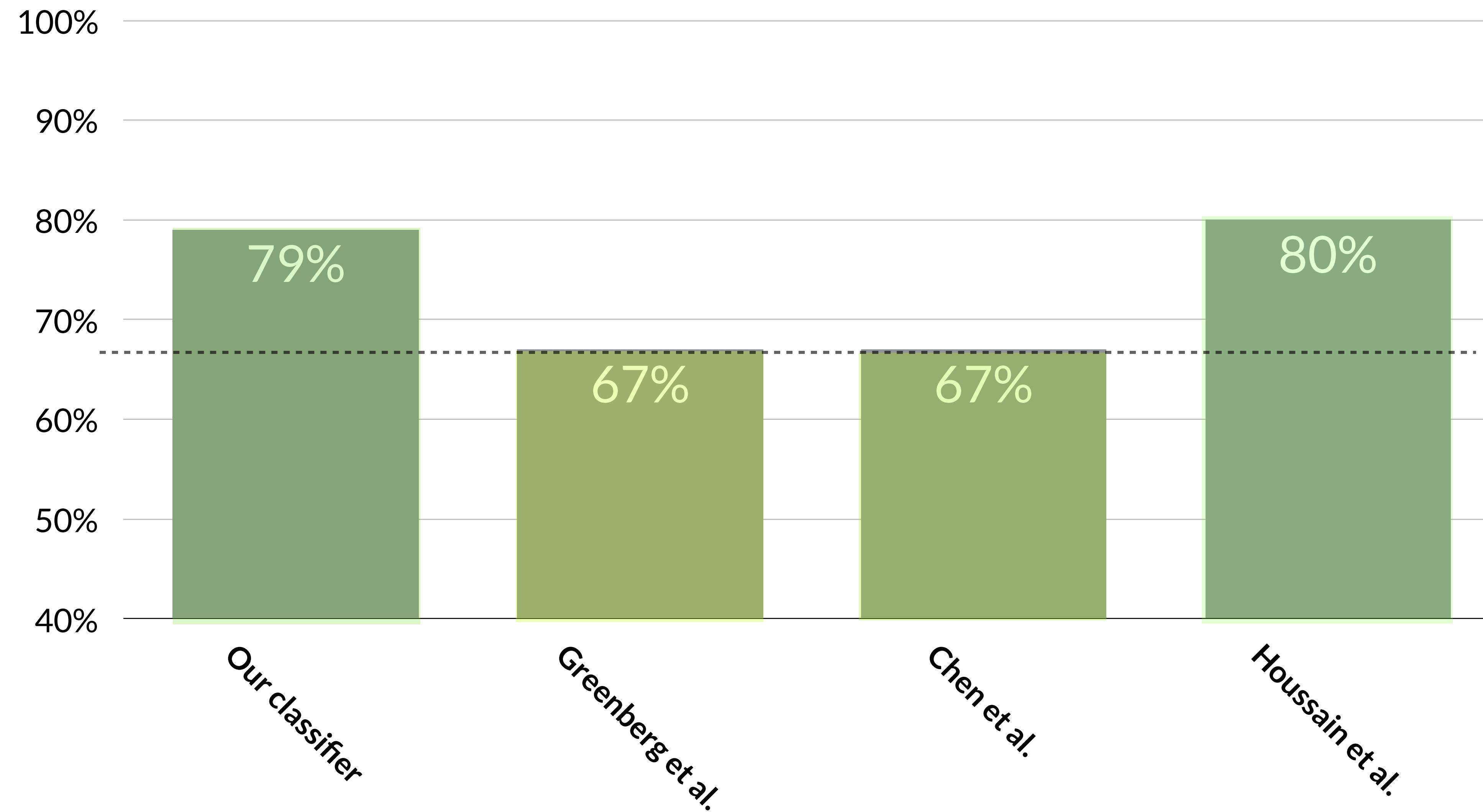
## Accuracy different classifiers (t=launch)





# Relative Classifier Performance

## Upper limit accuracy classifier in different studies (t=launch)





# Key Takeaways

---

1. *Discover unexamined attributes which contribute to campaign success*

- mean estimated delivery (direct)
- past success rate (direct)
- adoption speed (indirect)

2. *Build a classifier for campaign success*

- 78.8% accuracy ( $t = \text{launch}$ )
- 93.6% accuracy ( $t = \text{campaign completion}$ )

# Q&A

Any questions or comments?

# Back-up slides

In case of some tricky questions...

# Future Research

- Abstract concepts (e.g. quality of pitch)
- Other perk characteristics (shipping zone, shipping costs)
- Scarcity of perks (i.e. quantity limits)
- Equity-based crowdfunding on Indiegogo

# Managerial Implications

Funding duration

Number of days  
30  
Up to 60 days, but we recommend 30 or fewer

End on date & time

Projects with shorter durations have higher success rates. You won't be able to adjust your duration after you launch.

Reward #1 ?  
0 backers

Title			
Pledge amount	€0		
Description			
<a href="#">+ Add an item</a>			
Estimated delivery	December	v	2022
Shipping details	Select an option		
Limit availability	<input type="checkbox"/> Enable reward limit		

[Duplicate reward](#) [Delete](#)

# Performance Classifiers (t=launch)

Data available at launch

Model	Accuracy	Recall	Precision
Logistic Regression	76.5 %	75.2 %	71.9 %
Support Vector Machine	75.7 %	61.6 %	77.0 %
K-Nearest Neighbors	72.6 %	63.4 %	69.8 %
Decision Tree	73.4 %	72.5 %	67.7 %
Random Forest	78.9 %	74.5 %	76.1 %
XGBoost	77.4 %	74.5 %	73.4 %

Table 5.2.3.1 - Validation on test set (20%) using 10-fold cross validation (t=launch)

# Performance Classifiers (t=campaign completion)

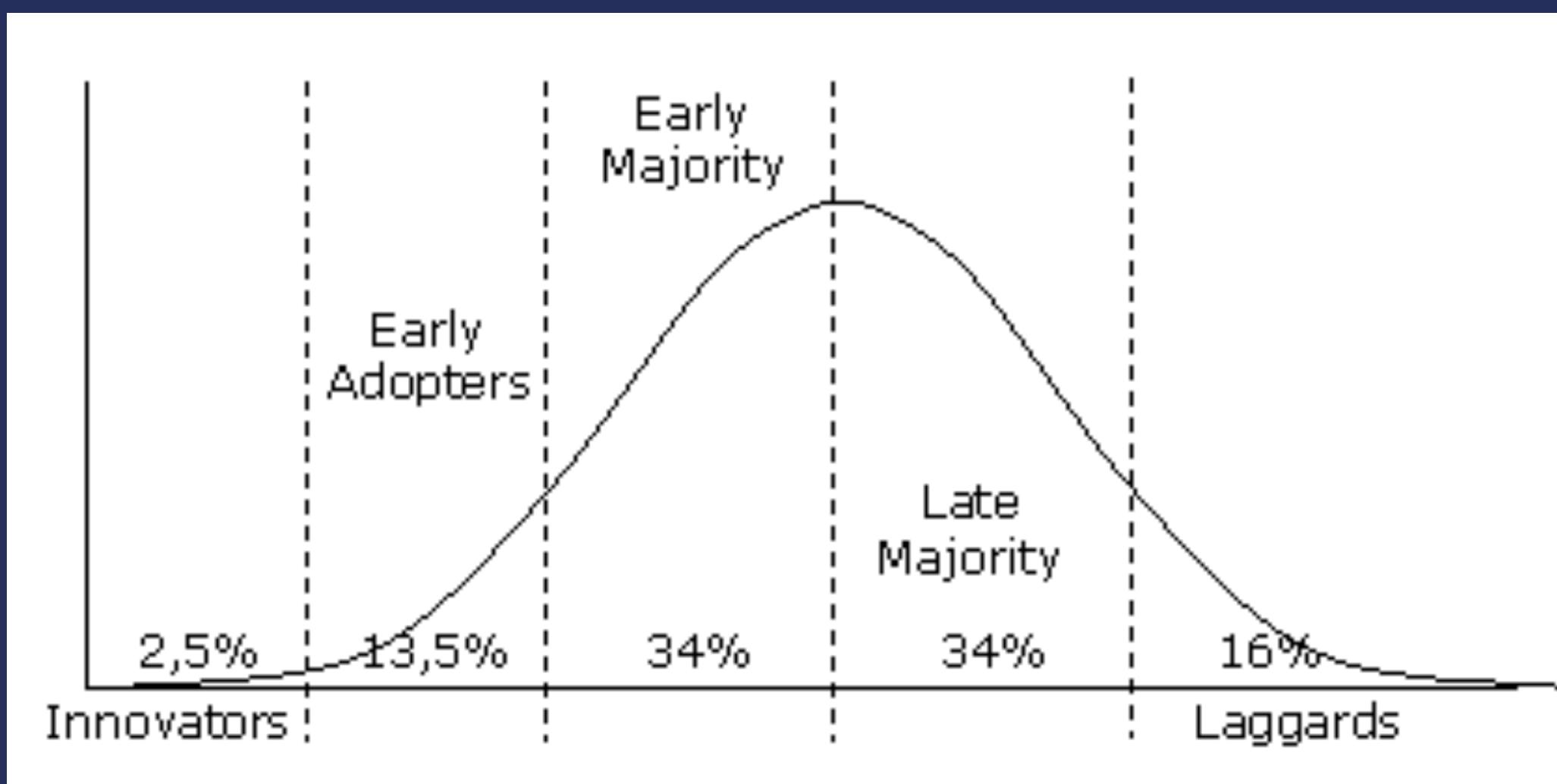
Data available at campaign completion

Model	Accuracy	Recall	Precision
Logistic Regression	92.7 %	90.2 %	92.6 %
Support Vector Machine	87.4 %	78.5 %	91.0 %
K-Nearest Neighbors	76.0 %	67.4 %	74.2 %
Decision Tree	91.3 %	90.3 %	89.6 %
Random Forest	92.7 %	92.8 %	90.5 %
XGBoost	93.6 %	90.4 %	95.3 %

Table 5.2.3.2 - Validation on test set (20%) using 10-fold cross validation (t=campaign completion)

# Diffusion Curve

In 2016 a study showed that 36% of all American consumers was not familiar with crowdfunding (Statista, n.d.). At that moment of time 241,873 unique creators had joined Kickstarter since 2009. Therefore, the total market size is estimated at  $241,873 / (1 - 0.36) = 377,927$ .



Year	#Projects*	#New creators	Percent of total	Adopter category
2009	874	874	0.2%	Innovators
2010	8,597	6,534	1.7%	Innovators
2011	23,262	17,679	4.7%	Early Adopters
2012	39,625	30,136	8.0%	Early Adopters
2013	44,094	33,511	8.9%	Early Majority
2014	66,558	50,584	13.4%	Early Majority
2015	76,867	58,419	15.5%	Early / Late Majority
2016	58,074	44,136	11.7%	Late Majority
2017	-	35,373**	9.4%	Late Majority
2018	-	28,753**	7.6%	Late Majority
2019	-	23,447**	6.2%	Late Majority / Laggards
2020	-	19,134**	5.1%	Laggards
2021	-	15,617**	4.1%	Laggards
2022	-	12,747**	3.4%	Laggards
TOT	317,951	376,943	≈ 100%	

# Predictors Success Rate

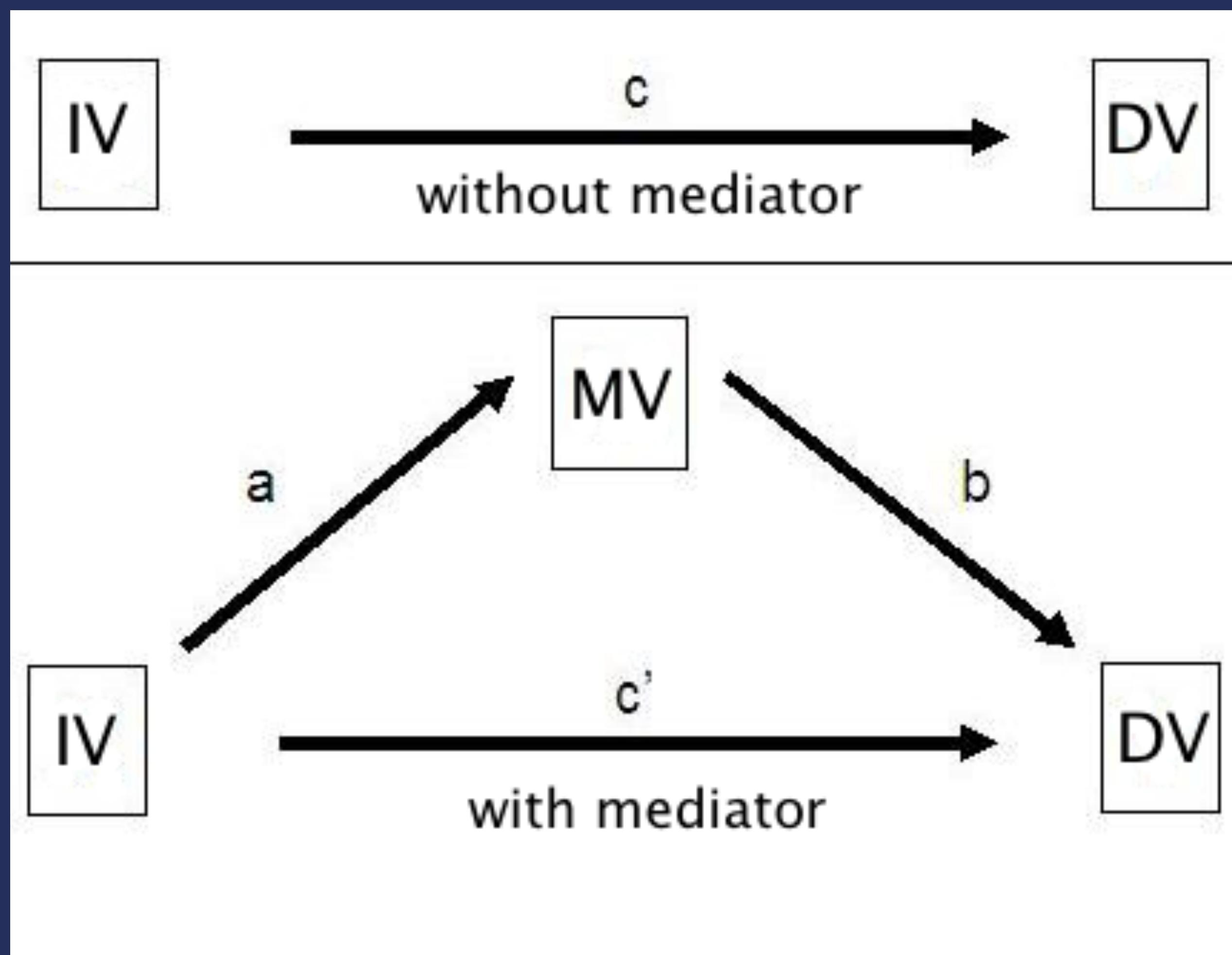
Attribute	Estimate	Std. Error	P-value
Goal	-5.58e-06	3.71e-06	0.13
Average number of backers required	-2.27e+01	3.02e+00	6.36e-14***
#Projects backed by creator	4.81e-01	1.71e-01	0.0049**
#Projects created by creator	4.07e-01	1.30e-01	0.0018**
#Facebook friends	1.33e-01	4.90e-02	0.0065**
Currency	9.92e-02	2.47e-02	5.96e-05***
Maximum pledge tier	-1.79e-01	6.02e-02	0.0029**
Number of rewards	5.60e-01	6.81e-02	<2e-16***
Number of collaborators	1.57e-01	6.31e-02	0.013*
Past success rate	4.53e-01	5.49e-02	<2e-16***
#Words in project description	1.67e-01	6.46e-02	0.0010**
Presence of video	2.56e-01	4.88e-02	1.62e-07***
#Additional videos	1.34e-01	6.27e-02	0.03*
#Images	5.32e-01	7.06e-02	4.90e-14***
Duration	-1.93e-01	5.17e-02	0.00019***
Estimated delivery	-1.91e-02	6.06e-02	0.0016**
Adoption	2.26e-01	5.72e-02	7.88e-05***

Dependent variable: *project success*

Attribute	Estimate	Std. Error	P-value
Goal	-2.16e-04	1.89e-05	<2e-16***
Average number of backers required	-5.65e+01	8.79+00	3.10e-10***
#Projects backed by creator	-8.99e-02	1.40e-01	0.52
#Projects created by creator	3.50e-01	1.63e-01	0.032*
#Facebook friends	1.20e-01	6.93e-02	0.08
Currency	1.40e-01	3.70e-02	0.00016***
Maximum pledge tier	1.57e-01	1.12e-01	0.16
Number of rewards	1.62e-01	1.06e-01	0.13
Number of collaborators	1.13e-01	1.20e-01	0.35
Past success rate	3.22e-01	7.44e-02	1.51e-05***
#Words in project description	4.52e-02	1.07e-01	0.67
Presence of video	1.09e-01	6.90e-02	0.11
#Additional videos	-1.98e-01	1.09e-01	0.068
#Images	2.30e-02	1.15e-01	0.84
Duration	-9.45e-02	7.11e-02	0.18
Estimated delivery	-2.90e-01	9.41e-02	0.0021**
Adoption	-8.35e-02	8.10e-02	0.30
Number of updates	7.83e-01	1.40e-01	2.02e-08***
Number of comments	1.05e+00	2.21e+00	0.63
Total number of backers	3.61e+01	2.18e+00	<2e-16***

Dependent variable: *project success*

# Adoption Speed (without mediator)



# Limitations

Topic	Solution / Comment
Sample representativeness	Increase sample size
Very wide range for some attributes	Remove outliers (e.g. using IQR)
Moment of data extraction (deadline)	Collect data at campaign launch and completion
Projects We Love (representativeness)	Increase sample size
Assumption: constant currency rates	Make use of dynamic currency rates
Assumptions: Diffusion Curve Uncertainty	Validate assumptions in the coming years
Visibility of Past Success Rate	Role of indirect and direct learnings from previous experience
Alternative ways to stack the models	Outside scope of BEP (according to Madis)
Analyses solely rely on scraped data	Take into account abstract concepts

# Alternative Deliverable

## 1 Average Number of Backers Required

Campaign Goal

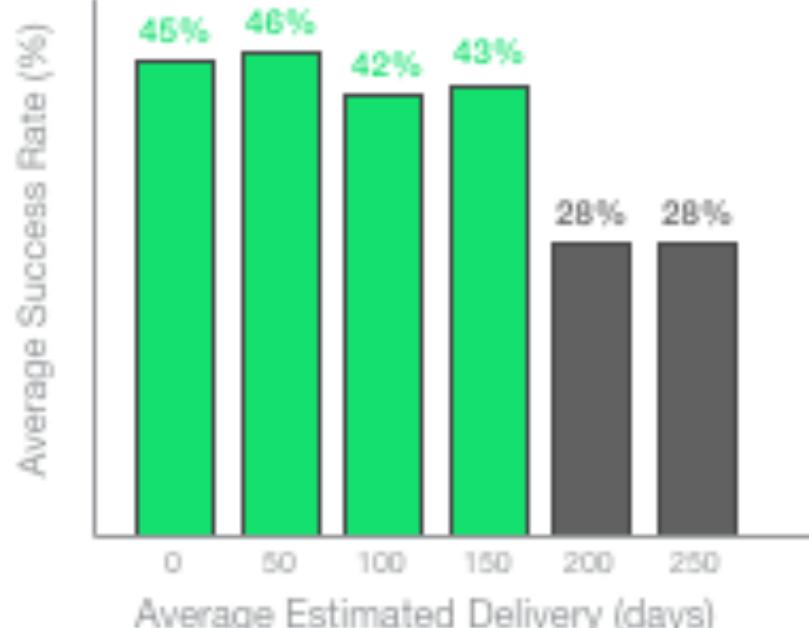
Avg. perk price

Simply divide the goal amount by the average perk value to get a rough estimate of how many backers you need. For most successful campaigns this figure is well below 500.

## 2 Estimated Delivery

The average shipping time is 95 days, so definitely no next-day delivery. However, make sure the average estimated delivery time of your perks is less than **half a year**.

Long shipping times? Ensure to also offer alternative perks with a faster estimated delivery date.



## 3 Number of Rewards



Failed campaigns: 6.1 rewards



Successful campaigns: 9.6 rewards

Successful campaigns typically offer a wider range of rewards. One of those campaigns even offered 74 unique perks! Therefore make sure there is something worthwhile at every pricing level, even a \$1 dollar perk will do.

## 4 Campaign Duration

Generally speaking, shorter campaigns perform better. In any case, never choose a duration longer than **49 days** since the average success rate for these campaigns is only **21.7%**. Moreover, the average total amount pledged is **58%** lower than for campaigns with a duration between 40 and 49 days.



Frequency distribution of campaign duration (successful campaigns only)

## 5 Number of Updates



Successful campaigns publish **ON AVERAGE FIVE TIMES** as many updates as failed ones

Launching your campaign does not mean the work is finished. Keep backers engaged through interesting and shareable (progress) updates, introduce additional perks and encourage your community to spread the word!

## BONUS! Random & Fun Facts

- ✓ At launch campaign success can be predicted with **78.8%** accuracy using machine learning algorithms (Random Forest).
- ✓ Practice makes perfect: creators whose first attempt was successful are more than **2.5 times** as likely to succeed their second project as well.
- ✓ The size of your social network matters: creators of successful projects have on average **118** more Facebook friends.

# Optimal Cut-off Levels

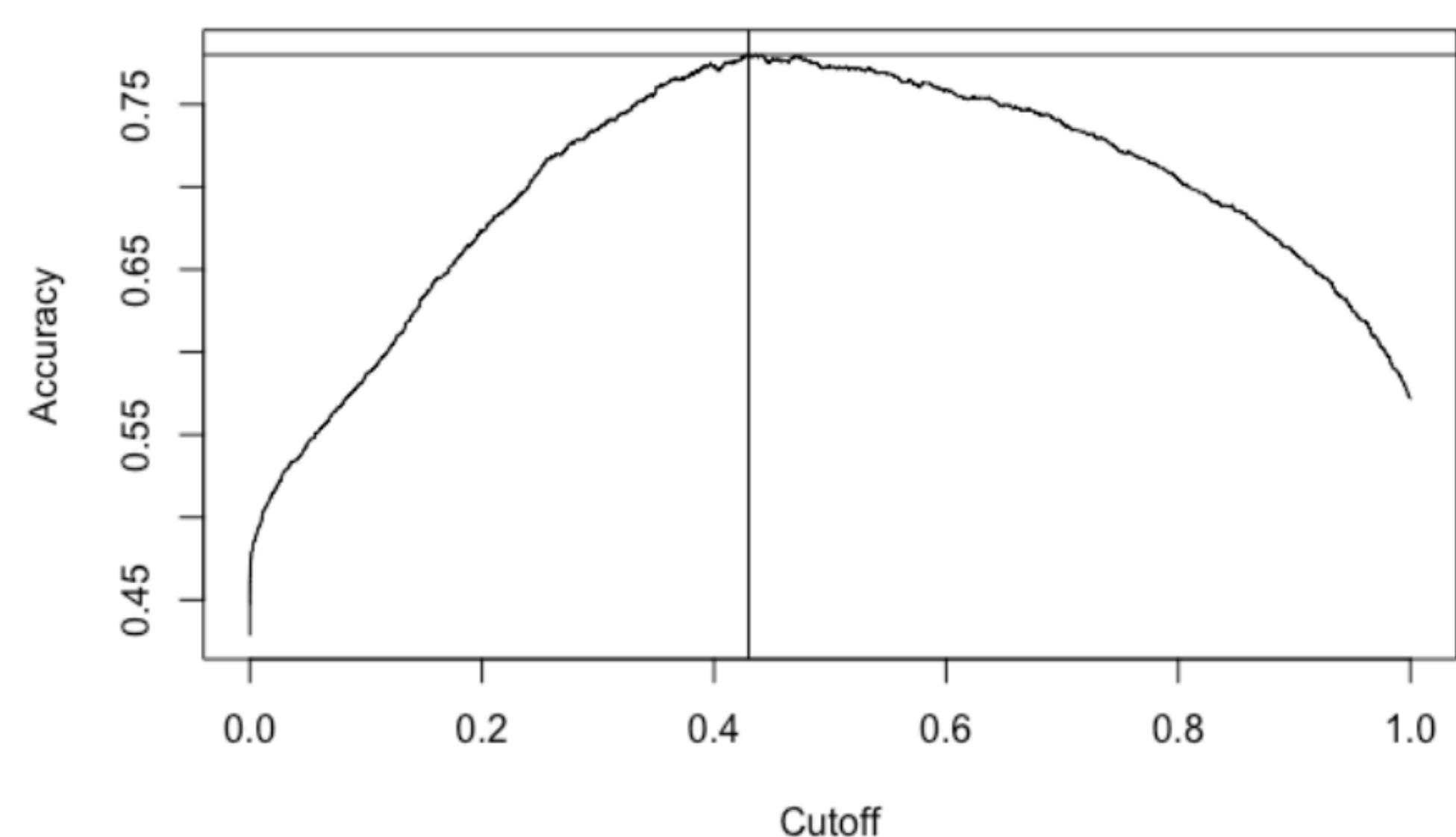


Figure X.I - Optimal cut-off level based on accuracy of model ( $t = \text{launch}$ ) - (0.430, 0.789)

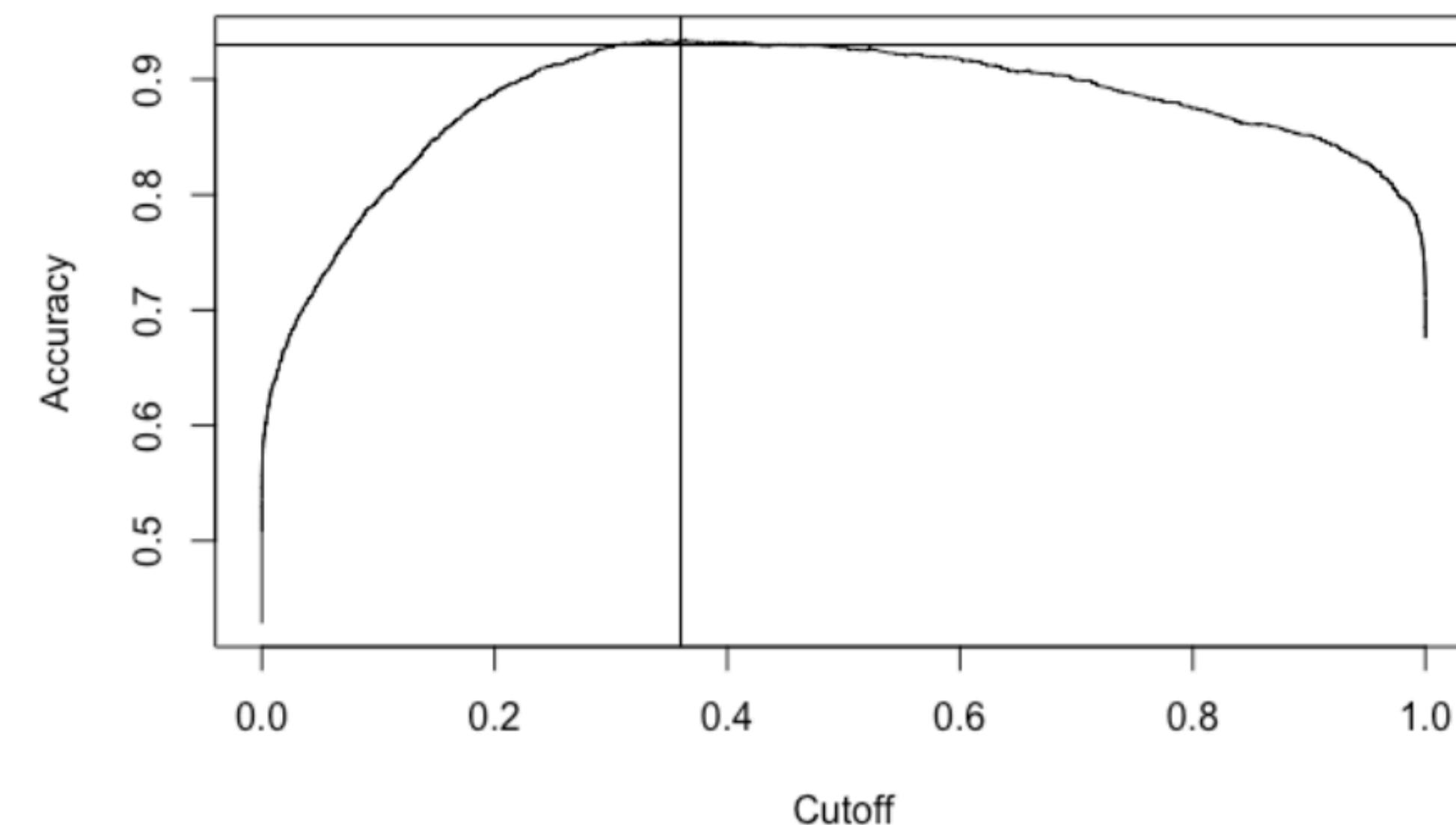


Figure X.II - Optimal cut-off level based on accuracy of model ( $t = \text{deadline}$ ) - (0.364, 0.936)

# ROC-Curves

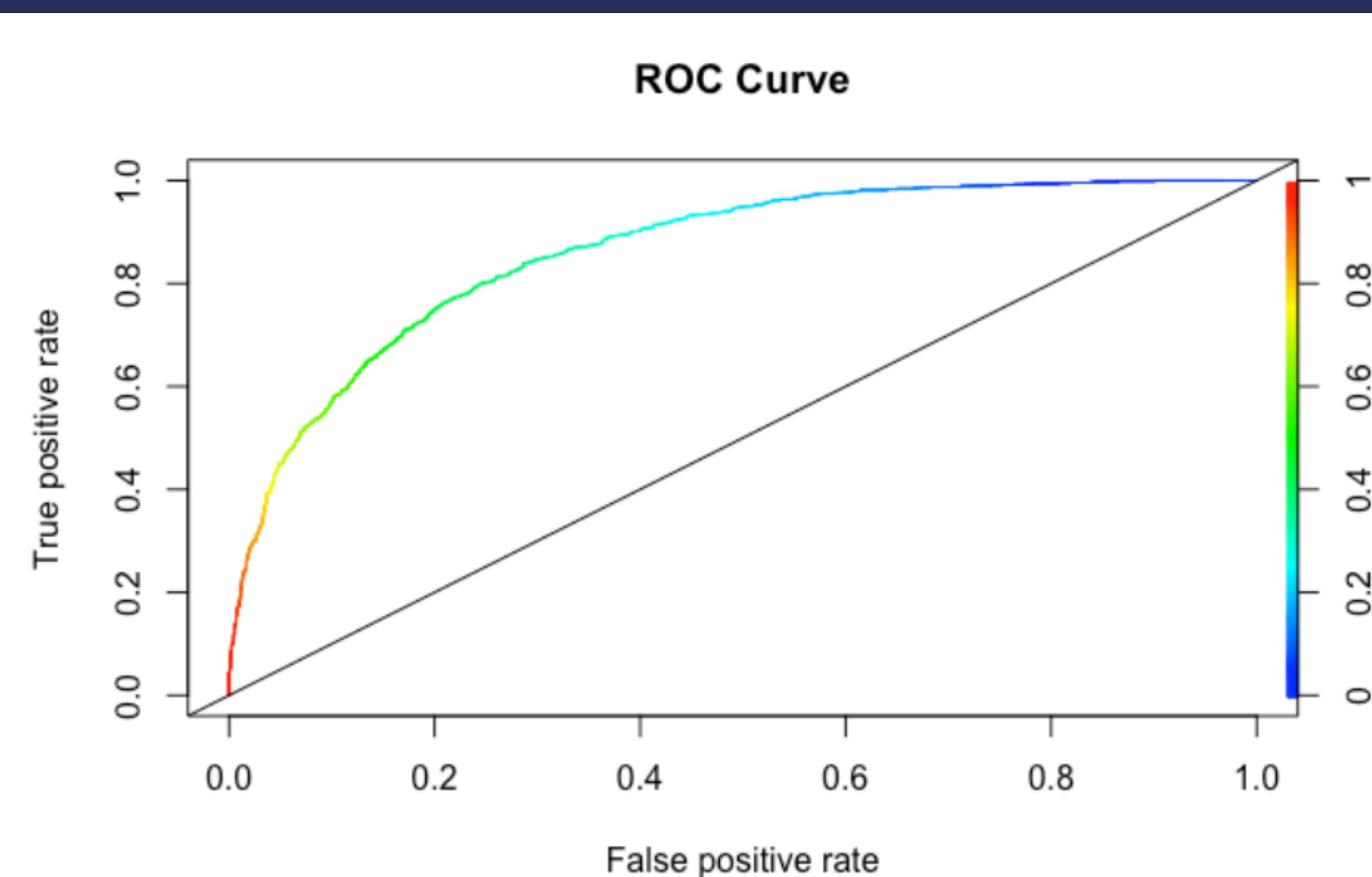


Figure XI.I - ROC-curve ( $t = \text{launch}$ ) - AUC = 0.861

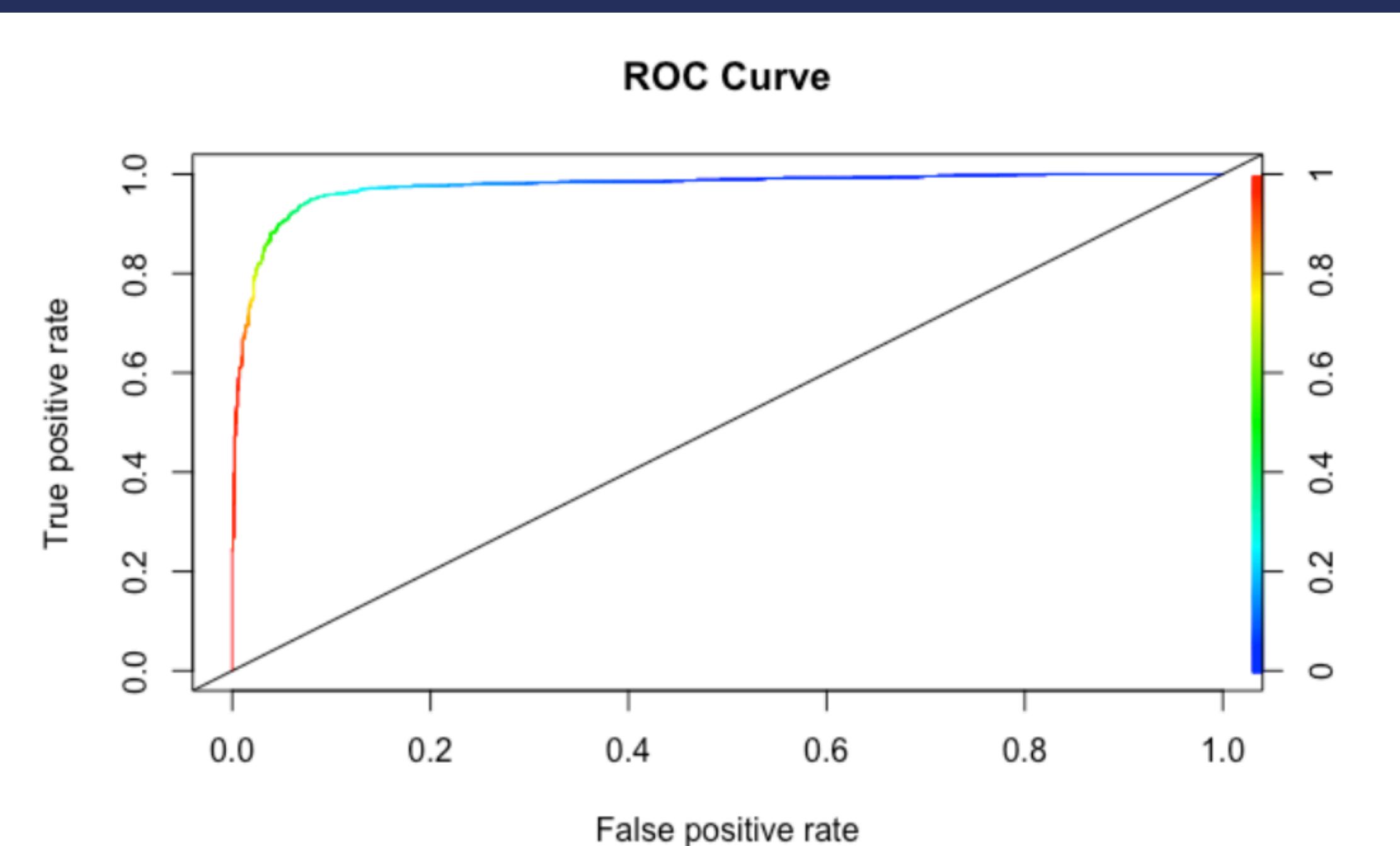


Figure XI.II - ROC-curve ( $t = \text{deadline}$ ) - AUC = 0.975

# Sample Representativeness

Category	Share of launched projects (2009-2017)	Share of launched projects (sample)	Average successful dollars pledged per project (2009-2017)	Average successful dollars pledged per project (sample)	Success Rate (2009-2017)	Success Rate (sample)
All	100%	100%	21450	32518	35.8%	41.1%
Games	8.9%	12.2%	54390	37851	34.7%	43.8%
Design	7.6%	9.5%	62258	76485	34.5%	45.5%
Technology	8.3%	11.5%	91996	147848	19.8%	26.3%
Film & Video	17.5%	11.6%	13740	10995	37.2%	41.4%
Music	14.6%	10.2%	6978	8469	49.7%	51.2%
Fashion	5.7%	7.0%	21058	47335	24.1%	31.1%
Publishing	10.5%	11.5%	8996	9018	30.4%	40.5%
Food	6.5%	6.4%	17083	16461	25.0%	30.2%
Art	7.4%	7.4%	6789	7428	40.6%	45.0%
Comics	2.7%	3.9%	11755	9149	53.0%	72.9%
Theater	2.9%	1.8%	5924	4322	60.2%	70.3%
Photography	2.9%	2.2%	9651	25779	30.5%	36.7%
Dance	1.0%	1.0%	5159	4510	62.5%	51.4%
Crafts	2.3%	2.8%	5403	2166	23.8%	28.7%
Journalism	1.3%	1.2%	10146	25803	21.7%	17.8%

# Pros and Cons Classifiers

Classification Model	Pros	Cons
Logistic Regression	Probabilistic approach, gives informations about statistical significance of features	The Logistic Regression Assumptions
K-NN	Simple to understand, fast and efficient	Need to choose the number of neighbours k
SVM	Performant, not biased by outliers, not sensitive to overfitting	Not appropriate for non linear problems, not the best choice for large number of features
Kernel SVM	High performance on nonlinear problems, not biased by outliers, not sensitive to overfitting	Not the best choice for large number of features, more complex
Naive Bayes	Efficient, not biased by outliers, works on nonlinear problems, probabilistic approach	Based on the assumption that features have same statistical relevance
Decision Tree Classification	Interpretability, no need for feature scaling, works on both linear / nonlinear problems	Poor results on too small datasets, overfitting can easily occur
Random Forest Classification	Powerful and accurate, good performance on many problems, including non linear	No interpretability, overfitting can easily occur, need to choose the number of trees

# Lessons Learned

- Be more **critical** about the **hypotheses argumentation**
- **Less is more** (number of hypotheses)
- Regression tables > Descriptive statistics
- Only start scraping data once you know what data **you need**
- **Technical Skills** (R / Machine Learning / Web Scraping)
- Overall a **good preparation** for my master in Data Science & Entrepreneurship!

# Thank you!

On to the next part of my data journey...

If any questions/remarks arise, please let me know!

Roy Klaasse Bos

r.j.klaasse.bos@student.tue.nl

# 1. Data collection

At your own risk..



# Scraping hub + Portia

Changes are saved automatically

PROJECT Show all projects  
Kickstarter\_final

SPIDER Show all spiders  
Successful\_Projects\_Gener...

SAMPLE PAGE Show all samples  
Basic and Advanced Break...

ITEMS

- 1 ABC campaign... text
- 1 ABC creator text
- 1 ABC totalNu... text
- 1 ABC numUpd... text
- 1 ABC numCo... text
- 1 ABC totalPled... text
- 1 ABC goal text
- 1 ABC location text
- 1 ABC category text
- 1 startD date
- 1 endD date

https://www.kickstarter.com/projects/125763333/basic-and-advanced-breakout-boards-plus-more

Explore Start a project About us Log in Sign up

## Basic and Advanced Breakout Boards Plus More!

Custom designed to be extra small. Gives you more space on your breadboard.

Environmentally friendly lead-free construction.

Created by

Frank Fox

124 backers pledged \$3,665 to help bring this project to life.

Inspector

No element selected

Extracted items 1 JSON

campaignTitle

Basic and Advanced Breakout Boards Plus More

category

DIY Electronics

creator

Frank Fox

endDate

2017-05-01 00:00:00

goal

\$500

location

High Point, NC

numComments

11



# Finetuning selectors

Annotation properties ×

**ANNOTATION**

Required ?

Selection mode

CSS selector

Selector

.pledge\_\_info > .pledge\_\_amount > .ple

Source

content

**FIELD**

Required ?

Vary ?

**EXTRACTORS** ▼

Pledge £2 or more

Digital Download

Receive a digital download of the single.

ESTIMATED DELIVERY

Sep 2017

Limited  
1 backer

```
graph LR; Selector[.pledge__info > .pledge__amount > .ple] -- "Red Arrow" --> Pledge["Pledge £2 or more"]; Selector -- "Dark Blue Arrow" --> Delivery["ESTIMATED DELIVERY"]
```



# Feed URL for spiders (Github Gist)

The screenshot shows a web browser window with the following details:

- Address bar: https://gist.githubusercontent.com/RoyKlaasseBos/36dcb5dcabc47d4f78223bfbe2c89c0d/raw
- Content area:

```
https://www.kickstarter.com/projects/1310584178/campus-food-ordering-app-foodorme?ref=ending_soon
https://www.kickstarter.com/projects/2048788161/new-level-streetwear-a-lifestyle-brand?ref=ending_soon
https://www.kickstarter.com/projects/149121530/media-wars-satirizing-news-as-it-happens?ref=ending_soon
https://www.kickstarter.com/projects/447184202/physician?ref=ending_soon
https://www.kickstarter.com/projects/1977420958/business-development-game-for-children?ref=ending_soon
https://www.kickstarter.com/projects/tomdifran/the-nostalgia-radio-project?ref=ending_soon
https://www.kickstarter.com/projects/1016208644/the-adventures-of-webzy?ref=ending_soon
https://www.kickstarter.com/projects/871605308/texas-collectibles-earn-and-care-store?ref=ending_soon
https://www.kickstarter.com/projects/867010291/kid-entrepreneur-finance-education-for-kids?ref=ending_soon
https://www.kickstarter.com/projects/1931396145/the-velveteen-rabbit-downloadable-mp3-audio-book?ref=ending_soon
```



# Scheduling

---

#1

31th of March  
(15:00)

#2

7th of April  
(15:00)

#3

14th of April  
(15:00)

#4

21th of April  
(15:00)

#5

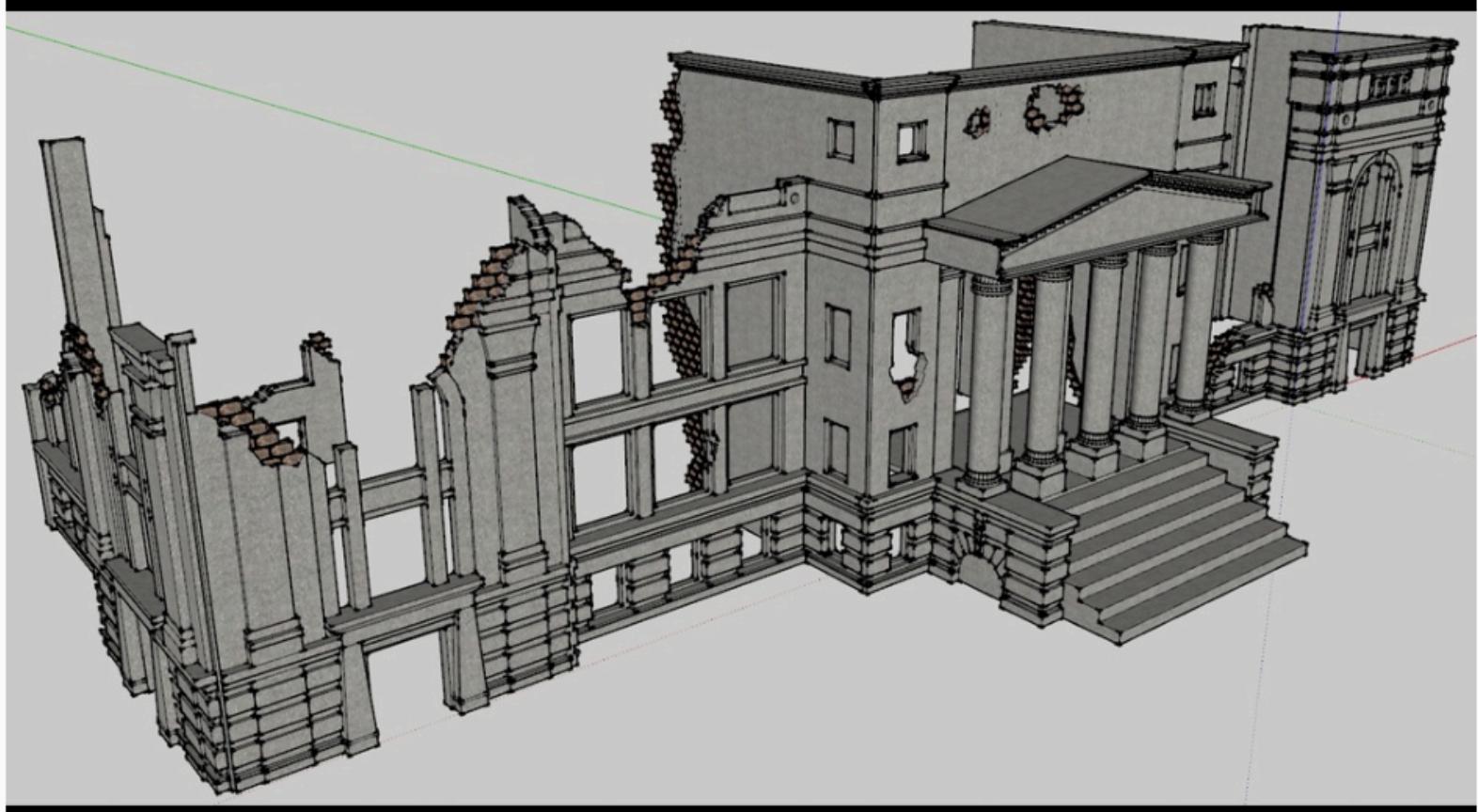
28th of April  
(15:00)

*29 days*



# Web scraping in a dynamic environment

5 minutes before deadline



**Ruined City Project**  
By Paul Deeming  
First created  
[Follow](#)

I want to create a ruined city for wargaming, this can be available in set scale cast resin or any scale stl file format.

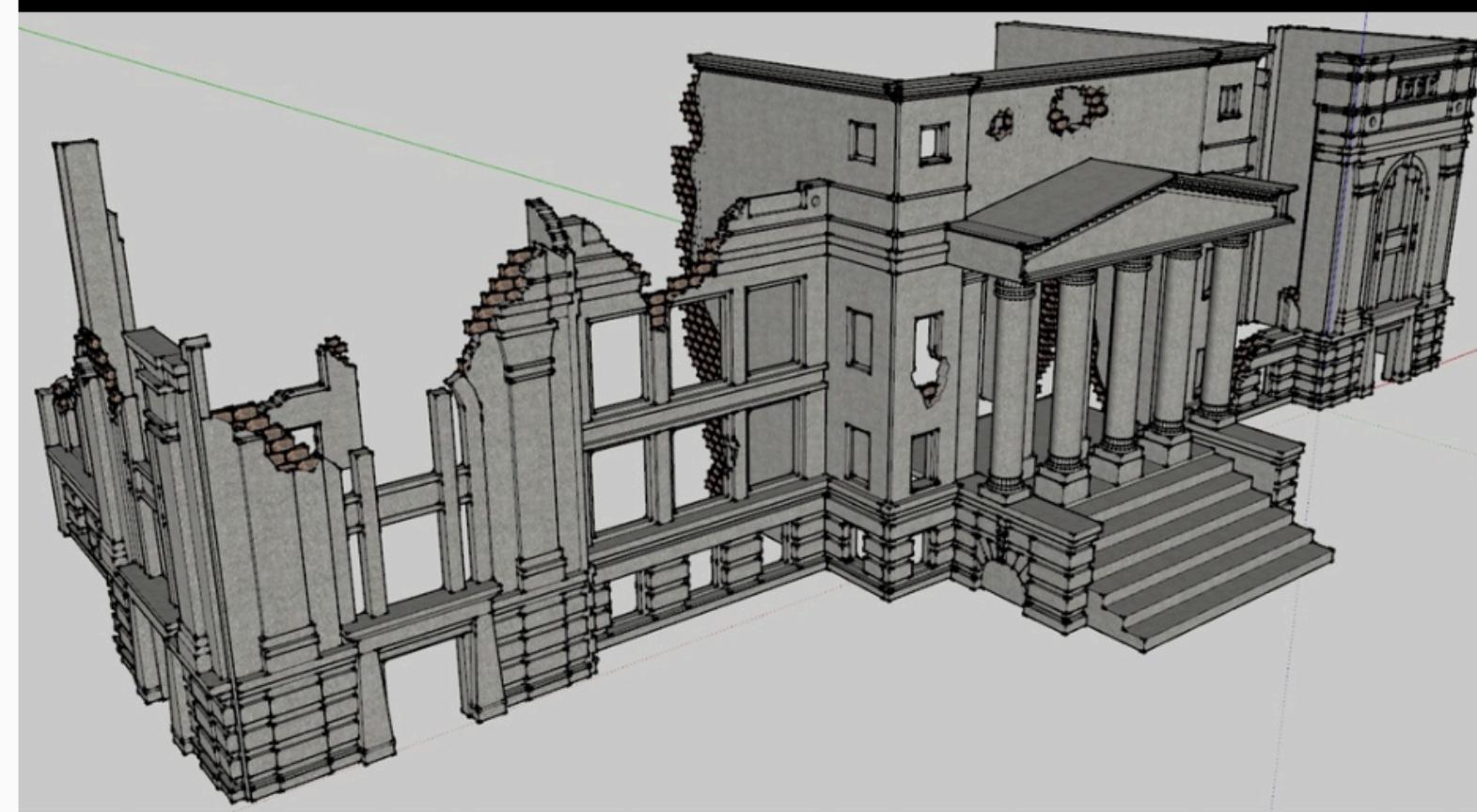
**£435**  
pledged of £1,500 goal  
**12**  
backers  
**5**  
minutes to go

[Back this project](#) [Save](#)

All or nothing. This project will only be funded if  
Tue, April 18 2017 2:14 PM CEST.

Tabletop Games Nottingham, UK

1 minute after deadline



**Ruined City Project**  
By Paul Deeming  
First created  
[Follow](#)

I want to create a ruined city for wargaming, this can be available in set scale cast resin or any scale stl file format.

**£435**  
pledged of £1,500 goal  
**12**  
backers  
**0**  
seconds to go

**Funding Unsuccessful**  
This project's funding goal was not reached less than a minute ago.

Tabletop Games Nottingham, UK



# Manual operations

## Configuring spiders



## Storing the data

[DATE]\_  
GeneralStats.xlsx

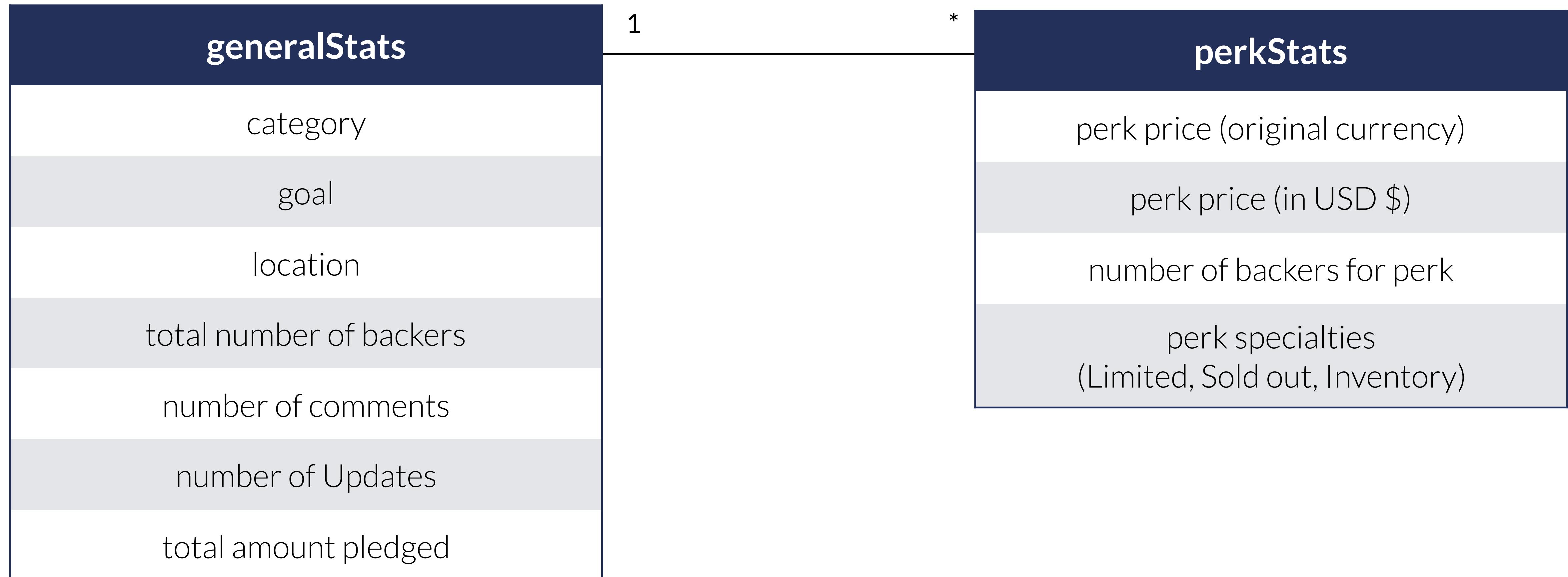
\_\_\_\_\_ Live campaigns  
\_\_\_\_\_ Successful campaigns  
\_\_\_\_\_ Unsuccessful campaigns

[DATE]\_  
PerkStats.xlsx

\_\_\_\_\_ Live campaigns  
\_\_\_\_\_ Successful campaigns  
\_\_\_\_\_ Unsuccessful campaigns



# Collected datasets



# 2. Data Preprocessing

From raw data to something more meaningful..



# Currency conversions

goal	originalValuta	typeof(goal)
â,-1,000		character
1,000	eur	character
1000	eur	double
1061	eur	double
1061	eur	double

## Code snippet:

```
convertValutaGoal <- function(valuta, exchangeRate) {  
  round(generalStats$goal[generalStats$valuta == valuta]  
  * exchangeRate, 2)  
}  
  
for(counter in 1:nrow(valutaMatrix)){  
  valutaShort = valutaMatrix[counter,3]  
  valutaExchange = as.double(valutaMatrix[counter,4])  
  
  generalStats$goal[generalStats$valuta == valutaShort]  
  <- convertValutaGoal(valutaShort, valutaExchange)  
}  
...
```

# Number of backers (total / for each perk)

Data output	Data type	
1,493 backers	(character)	
1 backer	(character)	
1,493 s	(character)	<b>Code snippet:</b> generalStats\$totalNumberBackers <- gsub(" backer", "", generalStats\$totalNumberBackers)
1	(character)	
1,493	(character)	generalStats\$totalNumberBackers <- gsub("s", "", generalStats\$totalNumberBackers)
1	(character)	
1493	(character)	generalStats\$totalNumberBackers <- gsub(", ", "", generalStats\$totalNumberBackers)
1	(character)	
1493	(integer)	generalStats\$totalNumberBackers <- as.integer(generalStats\$totalNumberBackers)
1	(integer)	



# Location

Location	City	CountryState
Austin, TX		
Oslo, Norway		
Austin, TX	Austin	TX
Oslo, Norway	Oslo	Norway
Austin, TX	Austin	Texas
Oslo, Norway	Oslo	Norway

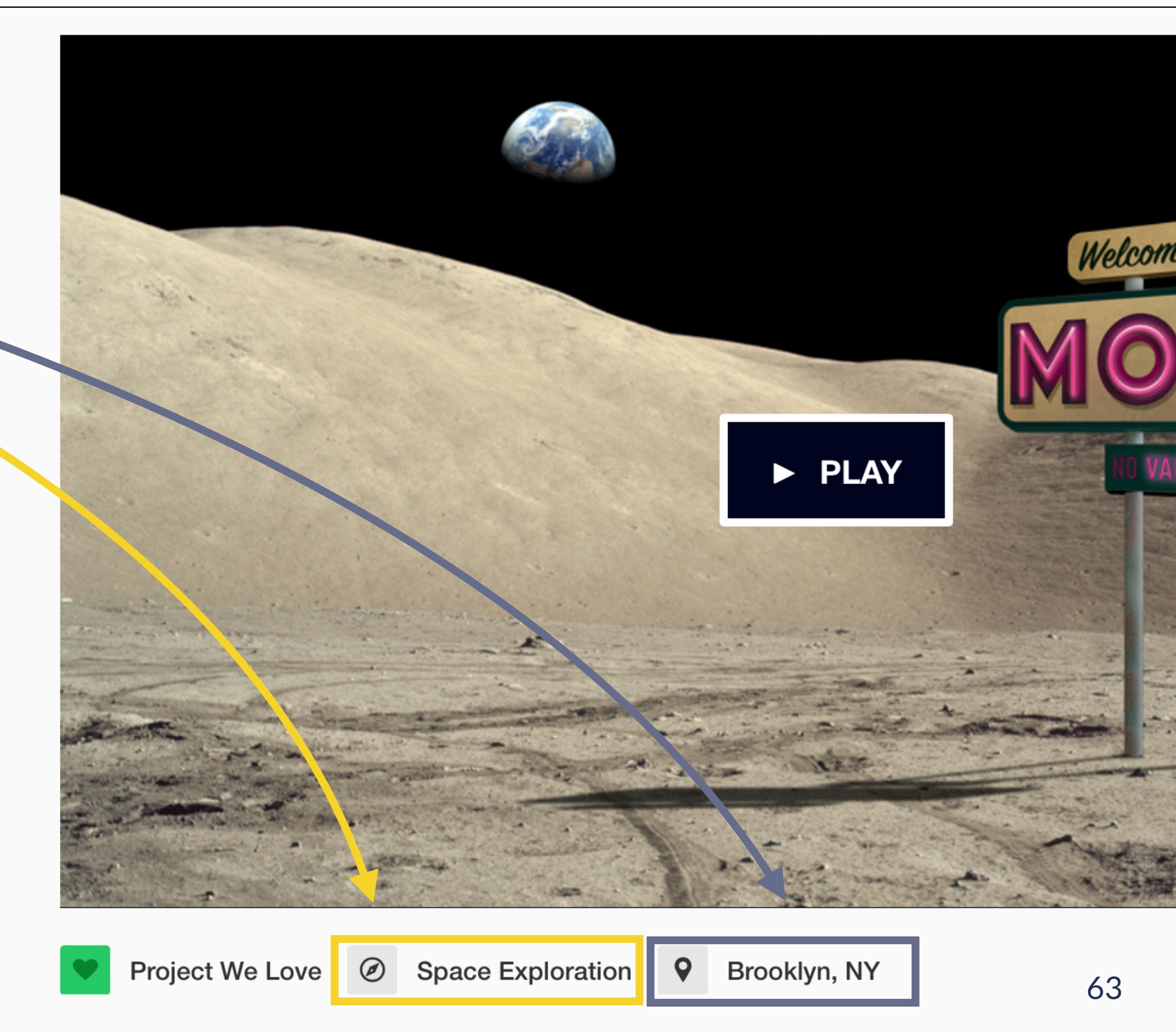
## Code snippet:

```
for(url in generalStats$url){  
  generalStats$commaLocation[generalStats$url == url] = gregexpr(pattern = "," ,  
  generalStats$location[generalStats$url == url])  
  [[1]]  
}  
...  
Campaign      Updates 1      Comments 0  
Austin, TX      Food
```



# Location (“Project We Love”)

Location	City	CountryState
Space Exploration		
Brooklyn, NY		
Brooklyn, NY	Brooklyn	NY
Brooklyn, NY	Brooklyn	New York





# “0-perks”

Pledge **NOK 0 or more**

## MULTIPLE REWARD PLEDGE/BACKING

Here is what you do:

1. Summarise the total amount ( in NOK, not \$USD) of all the items you want and insert the total amount here. Then go to check out.
2. Once you have pledged the amount you will get the chance to send me an email by clicking the "contact" creator button.
3. Send me the email with specified style name, size and colours.

We were informed by Kickstarter that is the way to solve this issue, and apologise for any inconvenience this may cause you.

Many Thanks for your pledge!

ESTIMATED DELIVERY

Oct 2017

SHIPS TO

Anywhere in the world

## Code snippet:

```
generalStats =
```

```
generalStats[generalStats$lowestPerk != 0, ]
```

# Canceled campaigns

## The First Photo Book App with Stickers, Video & Messages! **(Canceled)**

Create a photo book of friends & family events and everyone in the book can add messages, stickers & video.



\$2,105  
pledged of \$10,000 goal

40  
backers

### Funding Canceled

Funding for this project was canceled by the project creator on March 31.

### Code snippet:

```
generalStats =  
  generalStats[ !grepl("(Canceled)",  
  generalStats$campaignTitle), ]
```



# Hidden campaigns

---

## Sammich has been hidden for privacy.

This project has been removed from visibility at the request of the creator. It will remain permanently out of view.

Have a question? You can [ask the project creator](#).

Thanks for your patience.

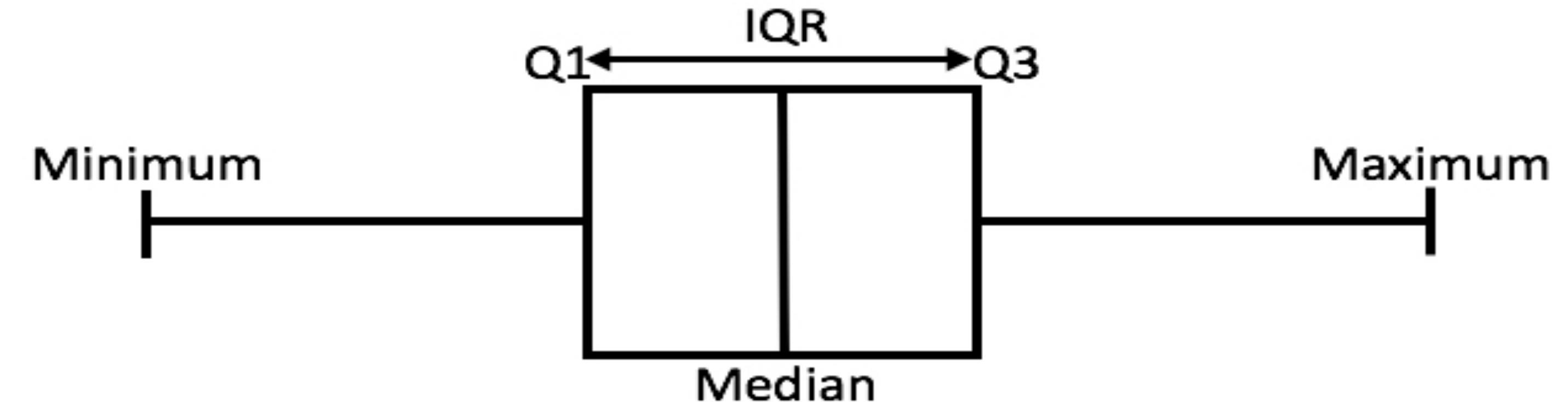
For example, [https://www.kickstarter.com/projects/1770343960/sammich?ref=ending\\_soon](https://www.kickstarter.com/projects/1770343960/sammich?ref=ending_soon)

### Code snippet:

```
generalStats = generalStats[generalStats$lowestPerk != "Inf", ]
```

# Remove outliers

Upper limit =  $Q3 + 1.5 * IQR$   
Lower limit =  $Q1 - 1.5 * IQR$



## Code snippet:

```
outliersHypothesisThree <- quantile(generalStats[, "lowestPerk"] )
```

```
hypothesisThreeLowerLimit <- max(outliersHypothesisThree[ 2 ] -  
1.5*(outliersHypothesisThree[ 4 ]-outliersHypothesisThree[ 2 ]),0)
```

```
hypothesisThreeUpperLimit <- max(outliersHypothesisThree[ 4 ] +  
1.5*(outliersHypothesisThree[ 4 ]-outliersHypothesisThree[ 2 ]),0)
```

```
generalStatsHypothesisThree <- generalStats[generalStats$lowestPerk >=  
hypothesisThreeLowerLimit & generalStats$lowestPerk <= hypothesisThreeUpperLimit, ]
```



# Examples of outliers (cheapest perk)

Perk includes	Cheapest perk	Percentage funding	URL
Be part of a documentary	MX\$ 150,000 (\$8100)	<1 %	<a href="https://www.kickstarter.com/projects/1048578123/documentary-the-vaquita">https://www.kickstarter.com/projects/1048578123/documentary-the-vaquita</a>
Hand made shoes	MX\$ 60,000 (\$3240)	0%	<a href="https://www.kickstarter.com/projects/489477071/miketi-sustainable-fashion?">https://www.kickstarter.com/projects/489477071/miketi-sustainable-fashion?</a>
Be a crew member during a flight	\$2,000	<1 %	<a href="https://www.kickstarter.com/projects/1401375406/the-douglas-dc-3-a-flying-">https://www.kickstarter.com/projects/1401375406/the-douglas-dc-3-a-flying-</a>
Translation of a book into English	\$10,000	2%	<a href="https://www.kickstarter.com/projects/1221178676/the-forwards-coverage-of-the-">https://www.kickstarter.com/projects/1221178676/the-forwards-coverage-of-the-</a>
Personal reference in every book sold	\$5,000	6%	<a href="https://www.kickstarter.com/projects/500297504/the-incredibles-scoobobell-and-">https://www.kickstarter.com/projects/500297504/the-incredibles-scoobobell-and-</a>
Free advertising for life (digital magazine)	\$3,000	0%	<a href="https://www.kickstarter.com/projects/656969276/catholic-life-international-">https://www.kickstarter.com/projects/656969276/catholic-life-international-</a>
UV-LED 3D Printer	HK\$ 12,624 (\$1624)	126%	<a href="https://www.kickstarter.com/projects/1037397144/d2k-plus-affordable-dlp-3d-">https://www.kickstarter.com/projects/1037397144/d2k-plus-affordable-dlp-3d-</a>
Luxury watch	HK\$ 5,990 (\$771)	67%	<a href="https://www.kickstarter.com/projects/phantomslab/futuristic-spectrum-swiss-made-">https://www.kickstarter.com/projects/phantomslab/futuristic-spectrum-swiss-made-</a>
Etched glass wall	\$1,000	101%	<a href="https://www.kickstarter.com/projects/181478271/little-red-brick-house?">https://www.kickstarter.com/projects/181478271/little-red-brick-house?</a>
Tailor-made app	\$1000	0%	<a href="https://www.kickstarter.com/projects/2045167832/shared-ride-food-delivery-">https://www.kickstarter.com/projects/2045167832/shared-ride-food-delivery-</a>
Sit-stand desk	\$555	168%	<a href="https://www.kickstarter.com/projects/1072402579/vivistand-next-generation-sit-">https://www.kickstarter.com/projects/1072402579/vivistand-next-generation-sit-</a>
Child's birthday party / t-shirts	\$1,000	<1 %	<a href="https://www.kickstarter.com/projects/1024575517/ninja-bounce-kids-">https://www.kickstarter.com/projects/1024575517/ninja-bounce-kids-</a>
2 books	\$1,000	<1 %	<a href="https://www.kickstarter.com/projects/1322327233/be-nice-be-happy-be-free-just-">https://www.kickstarter.com/projects/1322327233/be-nice-be-happy-be-free-just-</a>

# 3. Aggregated statistics

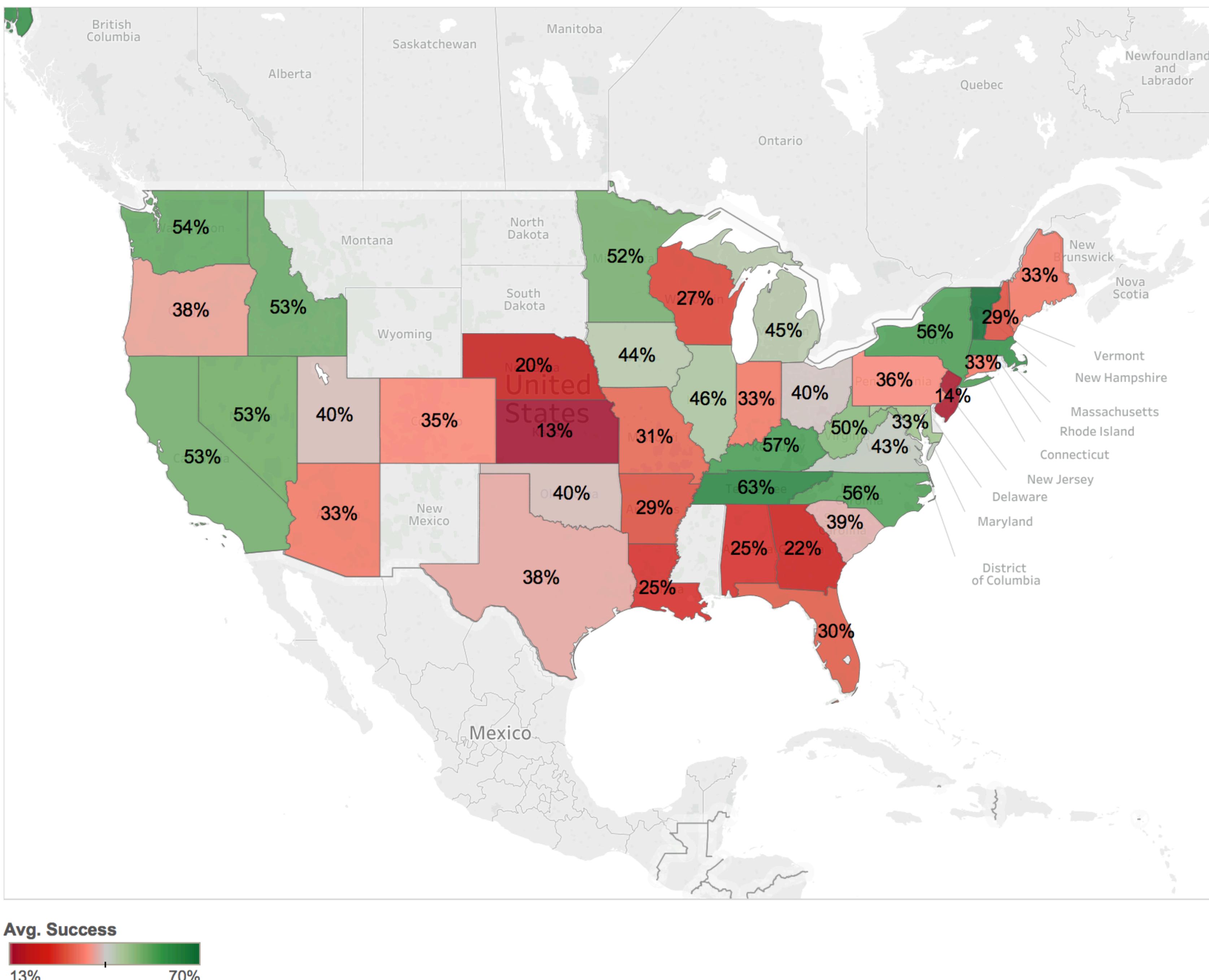
Just to get a better feel for the data



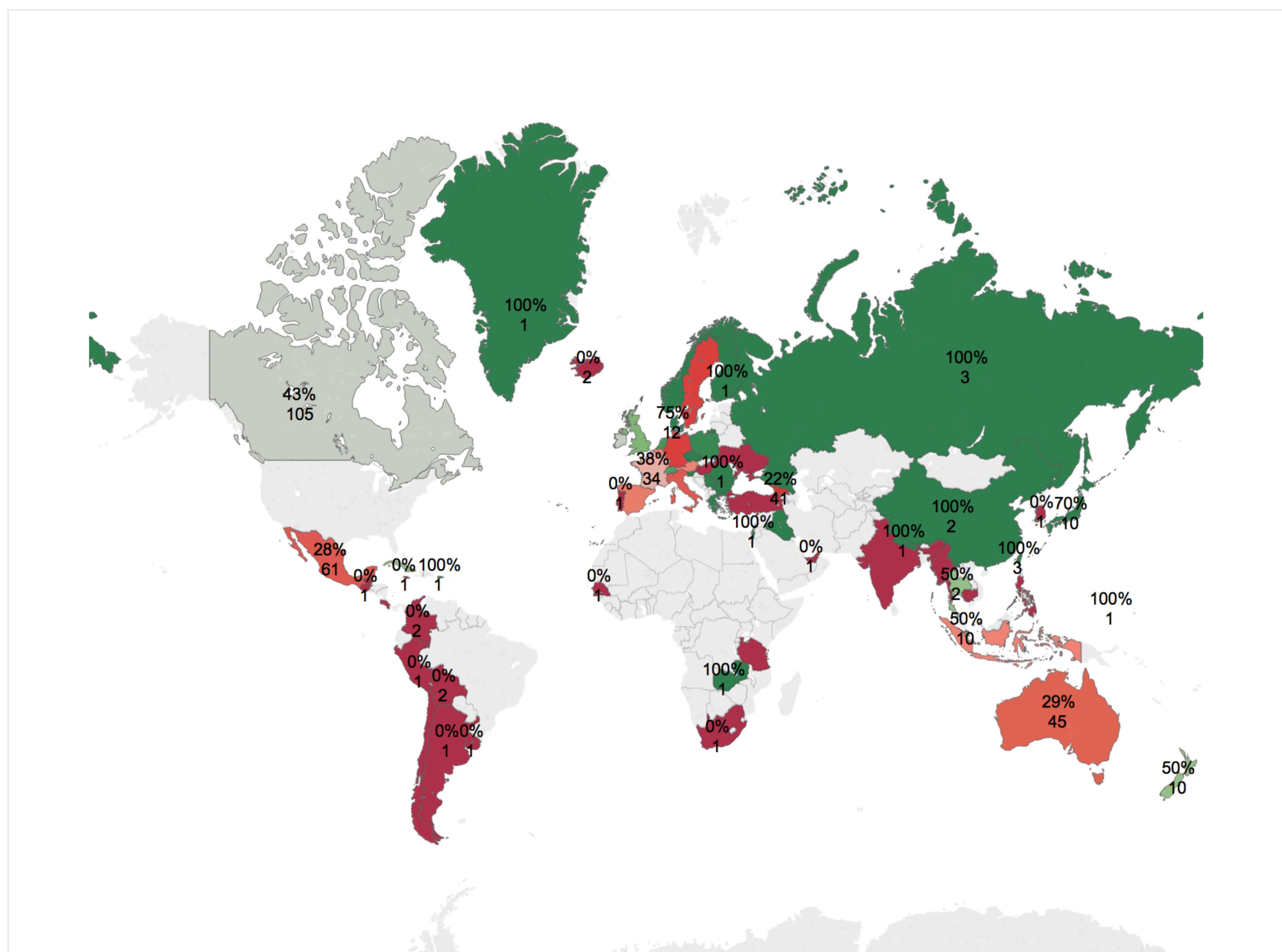
# Aggregated statistics (31-03 until 14-04)

Number of campaigns (n)	Total number of perks	Total pledge	Total number of backers	Average success rate campaigns
2185	16.268	~31M (\$)	297.557	44%

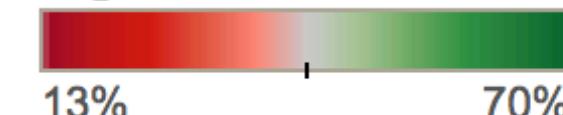
# Success rates of campaigns for states in USA



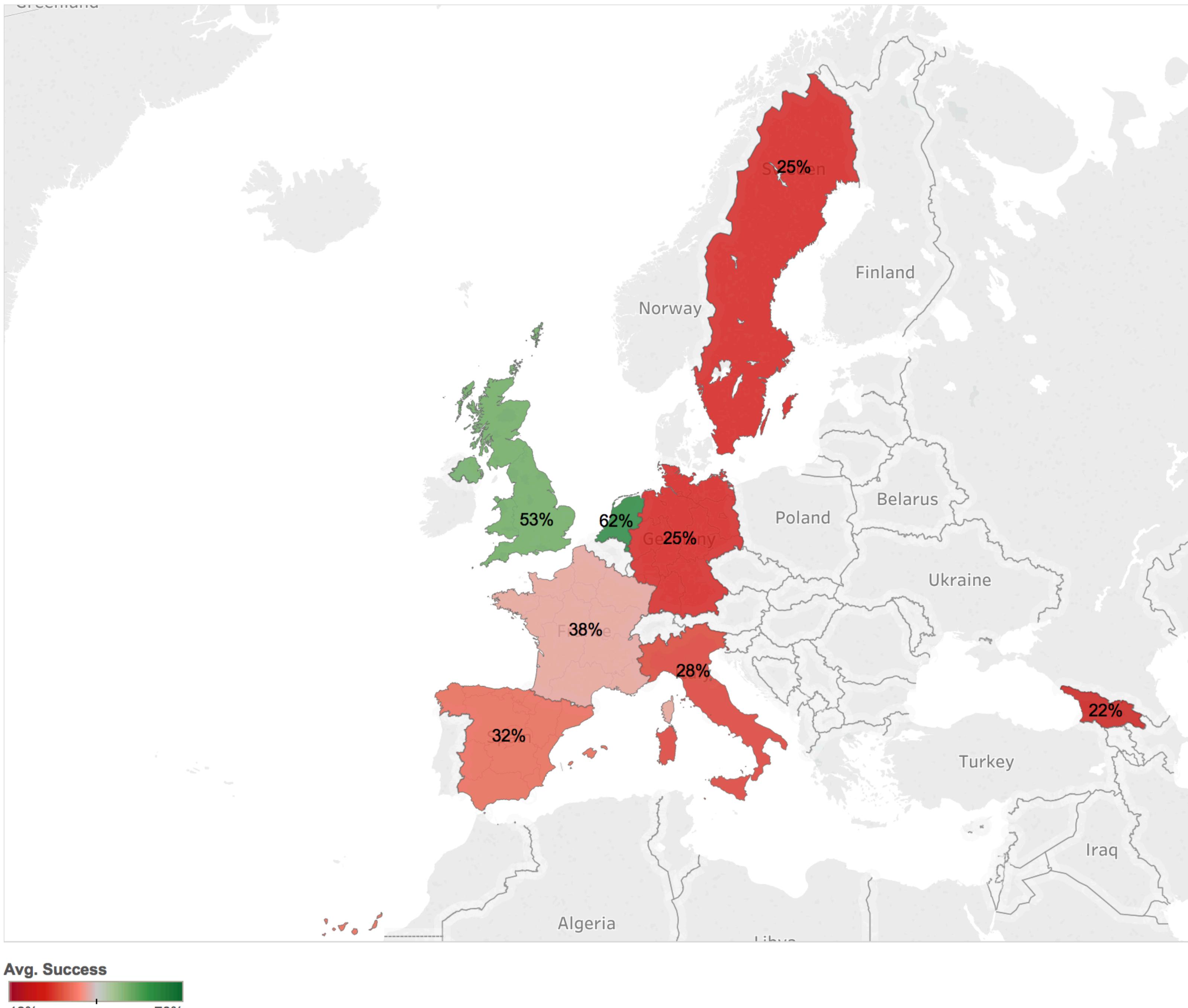
# Success rates of campaigns throughout the world



Avg. Success



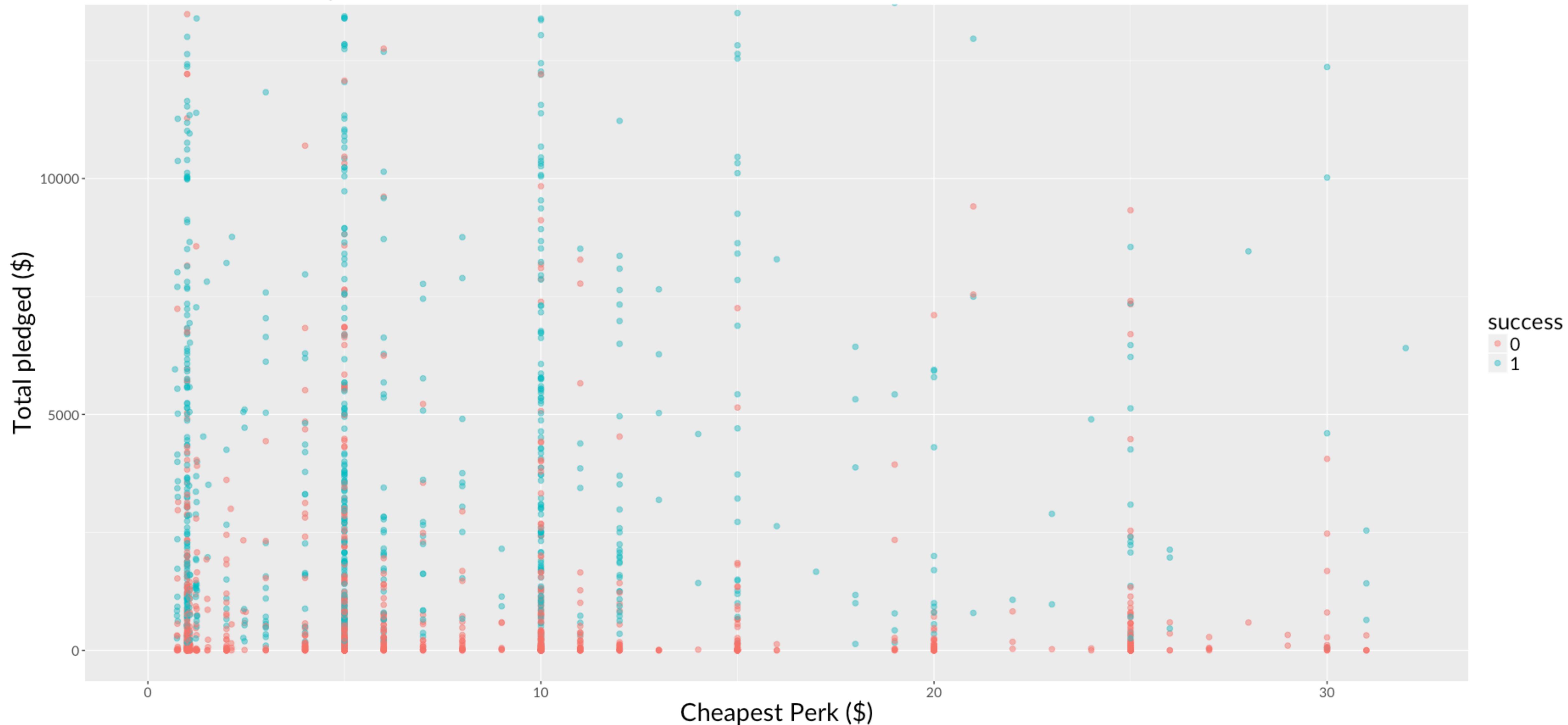
# Success rates of campaigns Europe (>=20 data points)



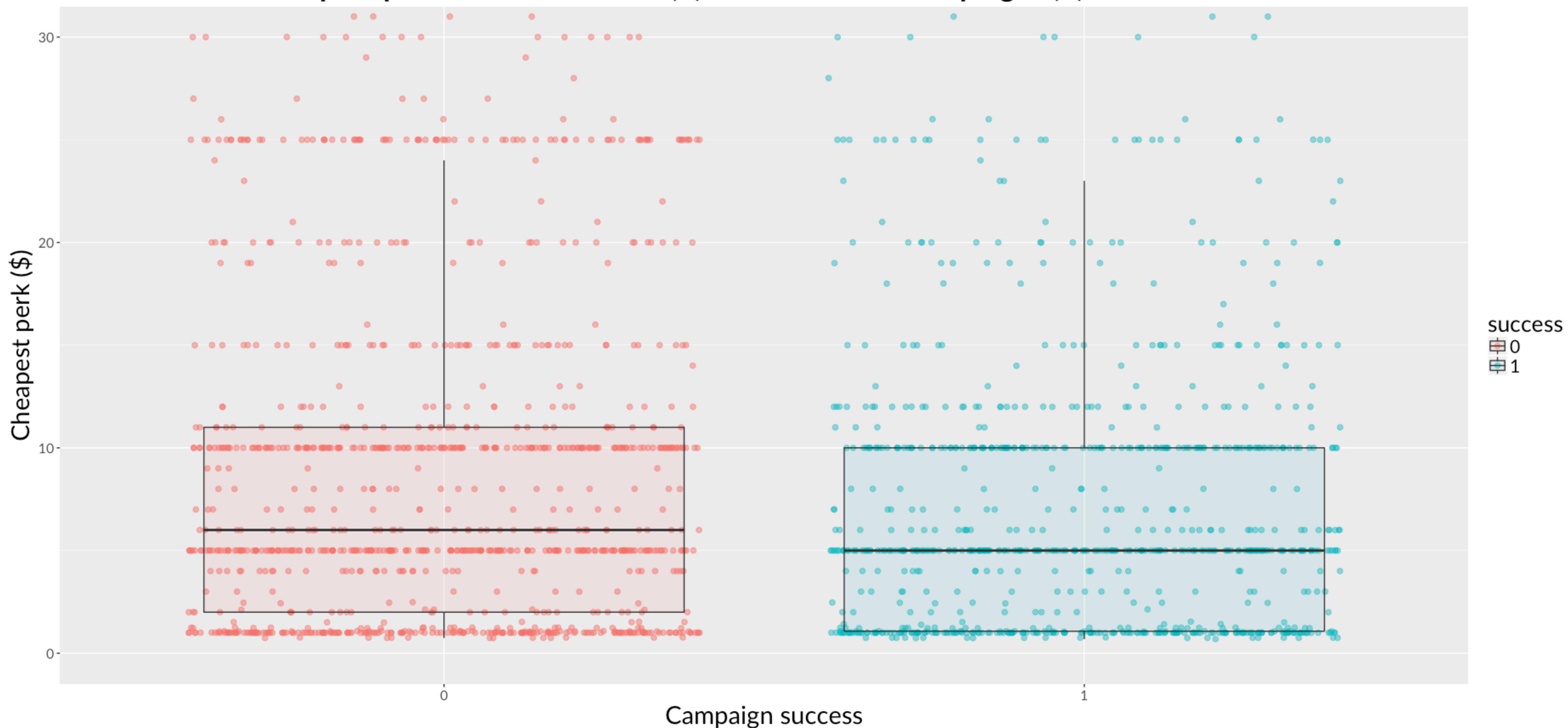
# 4. Visualisations

A picture is worth a thousand words

## Total amount pledged (\$) related to the cheapest perk (\$)



## Dollar value of cheapest perk for unsuccessful (0) and successful campaigns (1)





# Statistics cheapest perk

	Mean*	Standard deviation*	Median*	Mode*
Successful campaign	\$7.08	6.60	\$5	\$5
Unsuccessful campaign	\$8.65	7.73	\$6	\$5

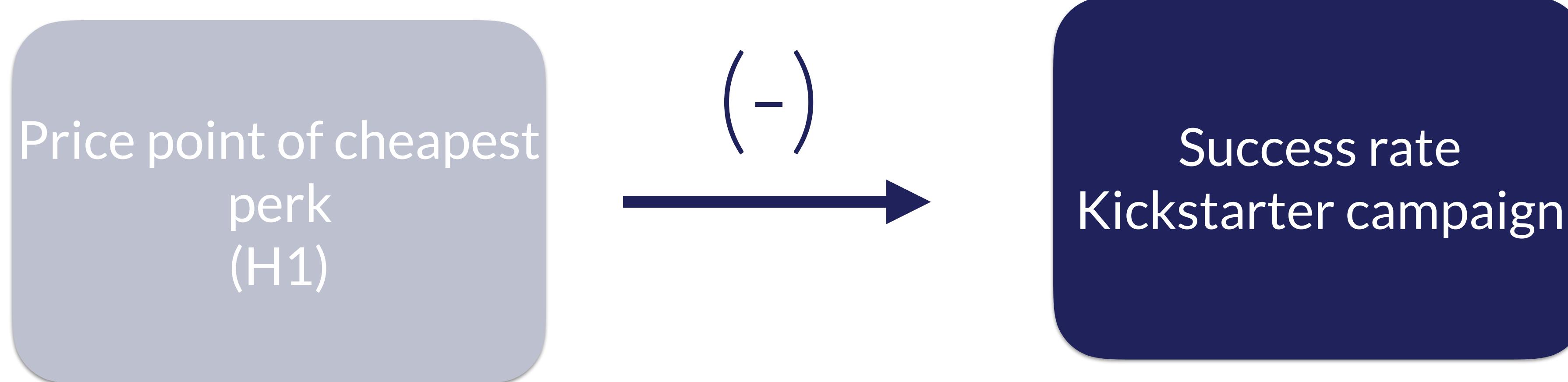
\*Only campaigns of which the cheapest perk fits within  $[Q1 - 1.5 * IQR, Q3 + 1.5 * IQR] = [0, 32]$  have been taken into account.



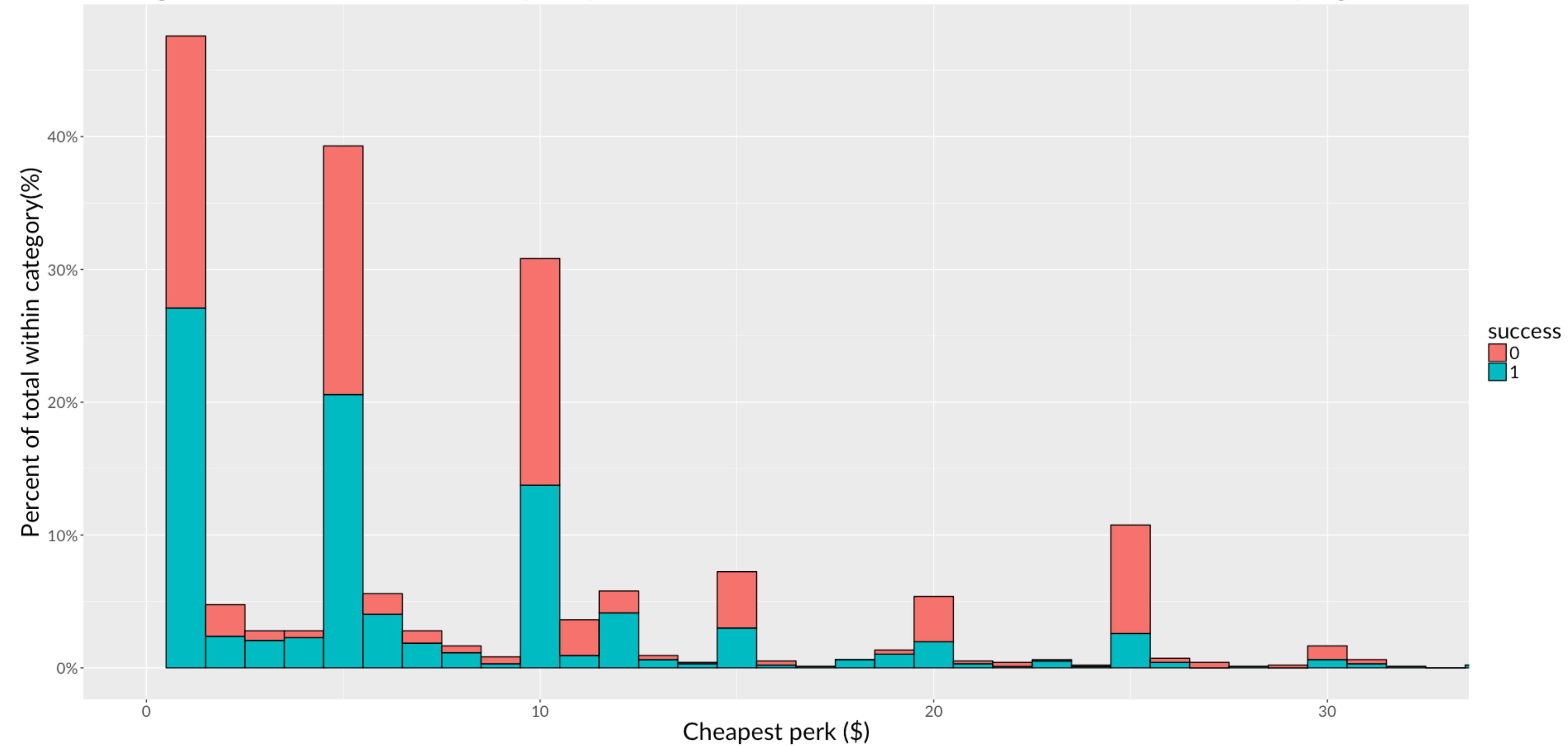
# Significant difference?

**Yes**, p-value: **1.67e-06 (\*\*\*)** - Analysis of Variance (ANOVA)

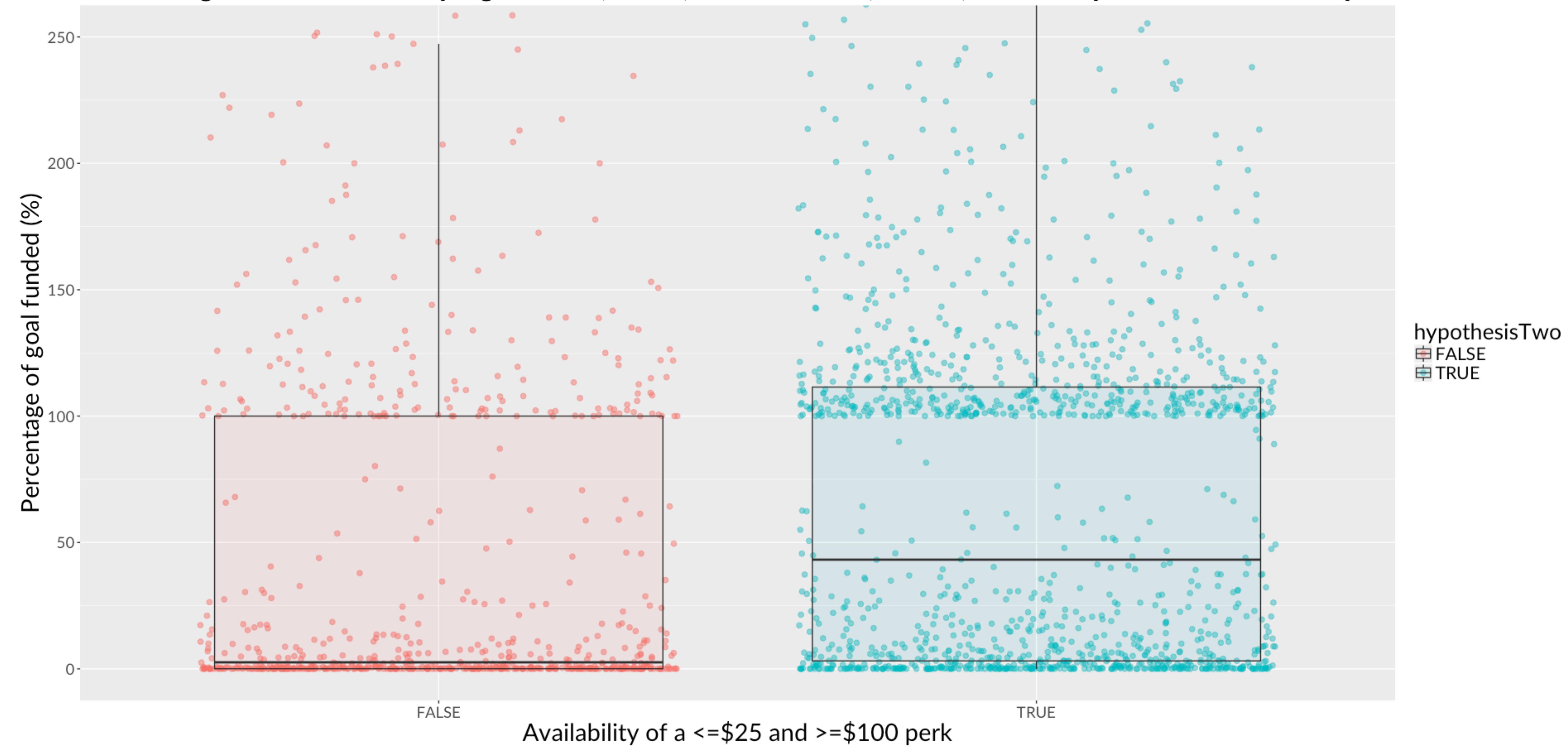
**Conclusion:** *The cheapest perk of unsuccessful campaigns is significantly higher (in terms of dollar value) than successful campaigns.*



# Histogram of distribution of cheapest perk levels for unsuccessful (0) and successful (1) campaigns stacked



## Percentage funded for campaigns with (TRUE) and without (FALSE) a <=\$25 perk AND >=\$100 perk



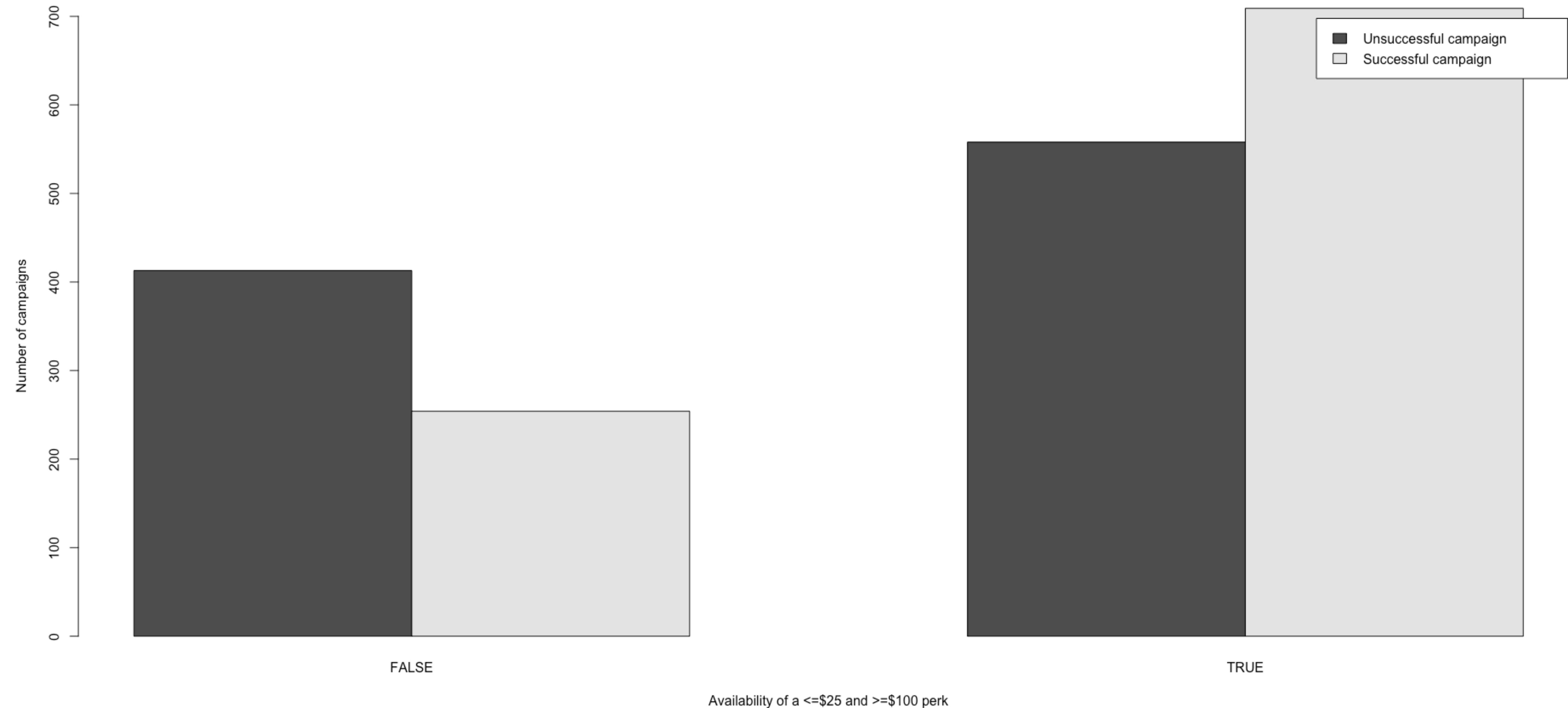


# Statistics percentage funded perk

	Mean percentageGoal*	Median percentageGoal*	Mean percentage of successful campaigns
With <= \$25 and >= \$100 perk	67.12%	43.20%	55.96%
Without <= \$25 and >= \$100 perk	40.92%	2.58%	38.08%

\*Only campaigns of which the percentageGoal is between [Q1-1.5\*IQR, Q3+1.5\*IQR] = [0, 285.53] have been taken into account.

**Success rate for campaigns with (TRUE) and without (FALSE) a  $\leq \$25$  perk AND  $\geq \$100$  perk**

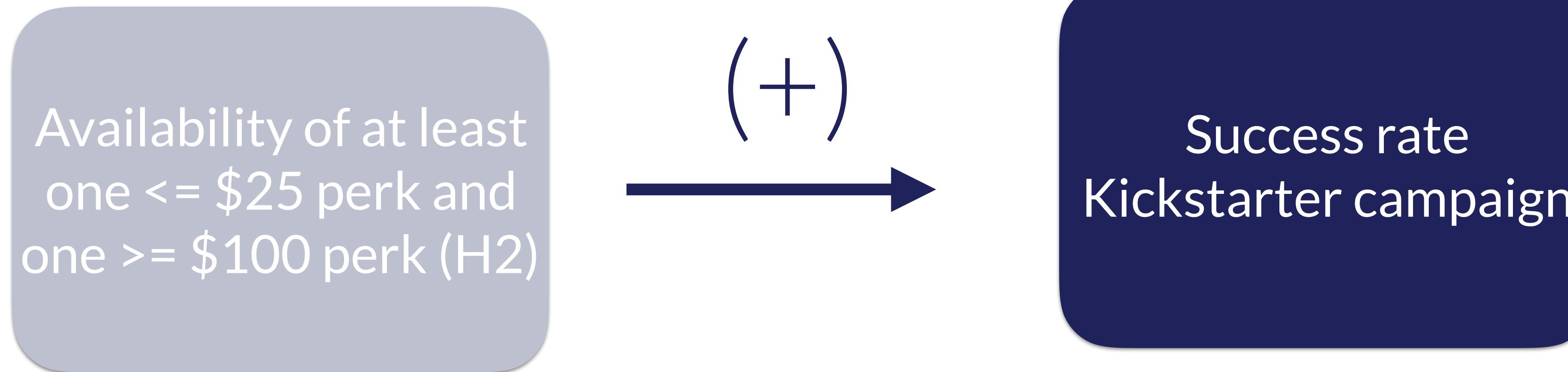




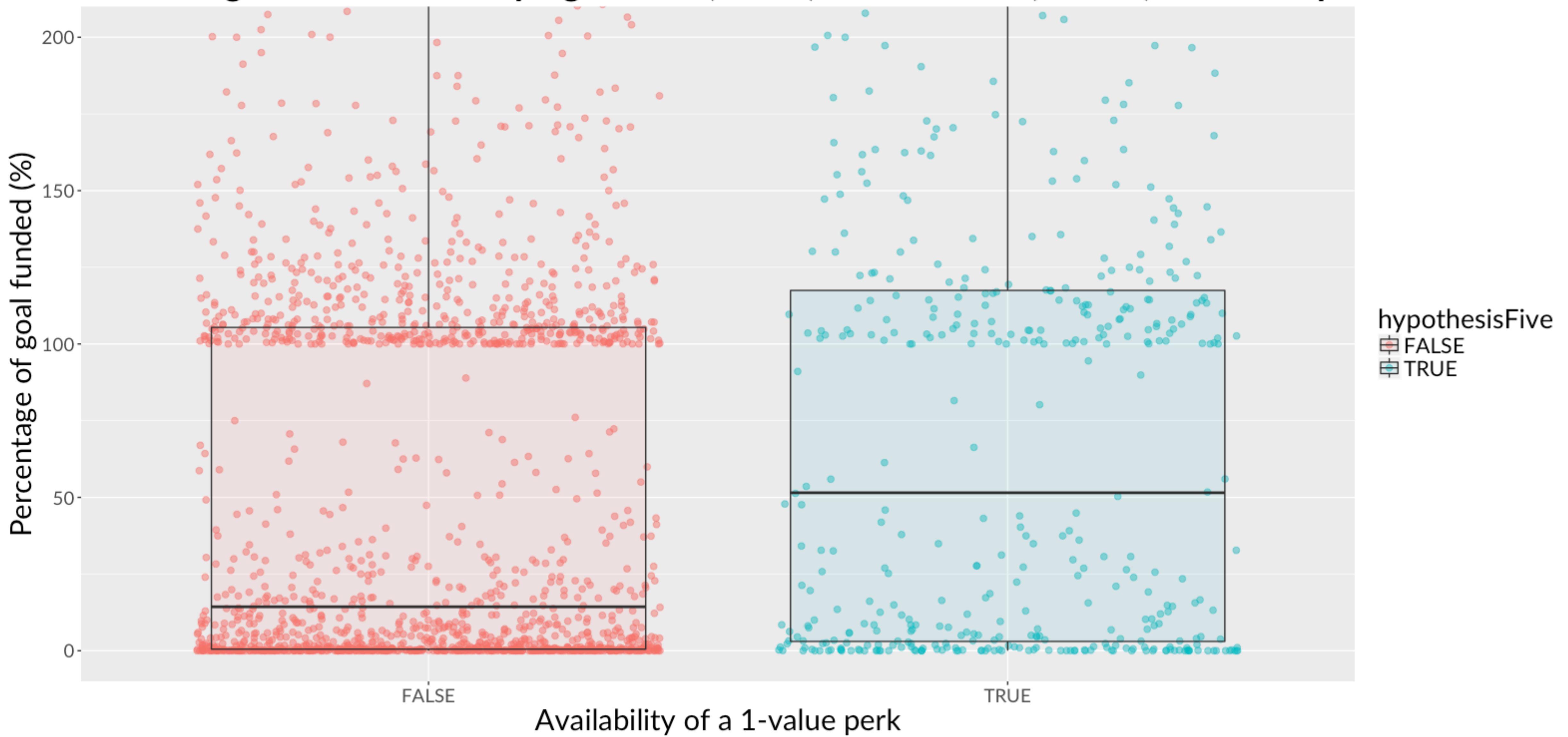
# Significant difference?

**Yes**, p-value: **1.115e-13 (\*\*\*)** - Chi square test

**Conclusion:** Campaigns with at least one  $\leq \$25$  perk and one  $\geq \$100$  perk are more likely to succeed than other campaigns (with one perk or campaigns that don't meet aforementioned criteria).



# Percentage funded for campaigns with (TRUE) and without (FALSE) a 1-value perk





# Statistics percentage funded

	Mean percentageGoal*	Median percentageGoal*	Mean percentage of successful campaigns *
With 1-value perk	73.27%	51.50%	55.70%
Without 1-value perk	53.58%	14.32%	48.02%

\*Only campaigns of which the percentageGoal is between  $[Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR] = [0, 285.53]$  have been taken into account.

**Success rate for campaigns with and without 1-value perk**

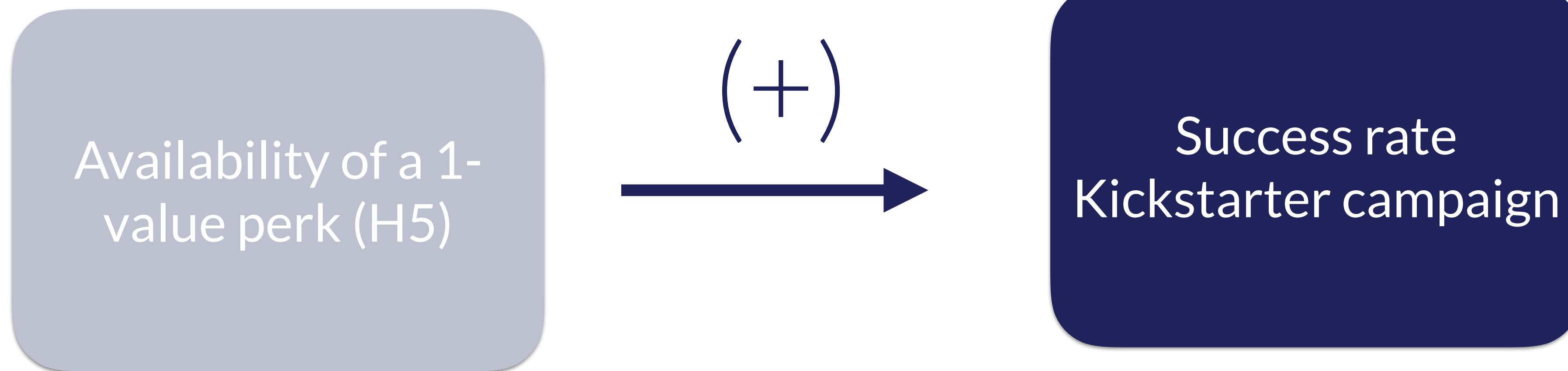




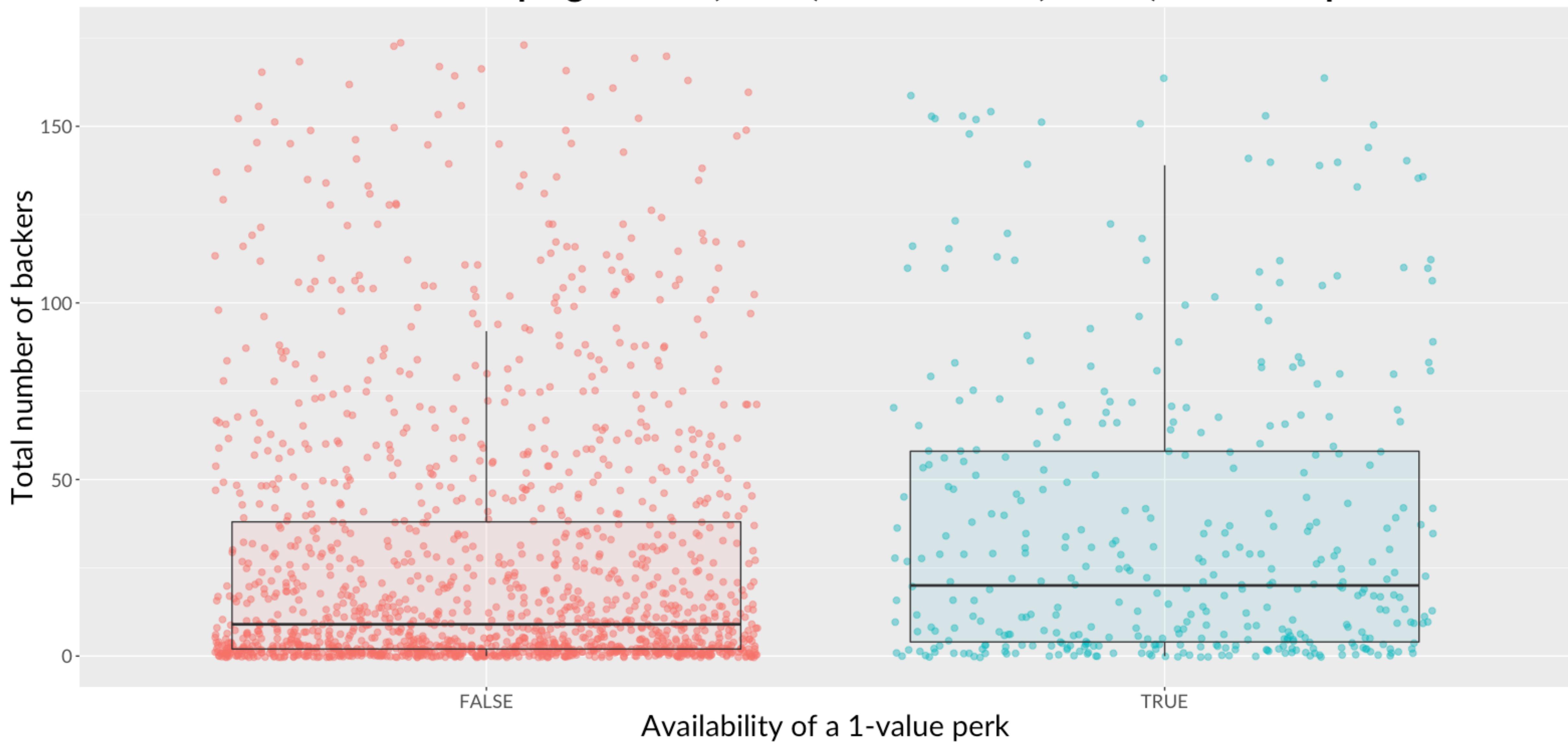
# Significant difference?

**Yes**, p-value: **0.005161 (\*\*)** - Chi square test

**Conclusion:** Campaigns with a 1-value perk are more likely to succeed than campaigns without one.



## Number of backers for campaigns with (TRUE) and without (FALSE) a 1-value perk





# Statistics total number of backers

	Mean total number of backers*	Median total number of backers*
With 1-value perk	37	31
Without 1-value perk	27	13

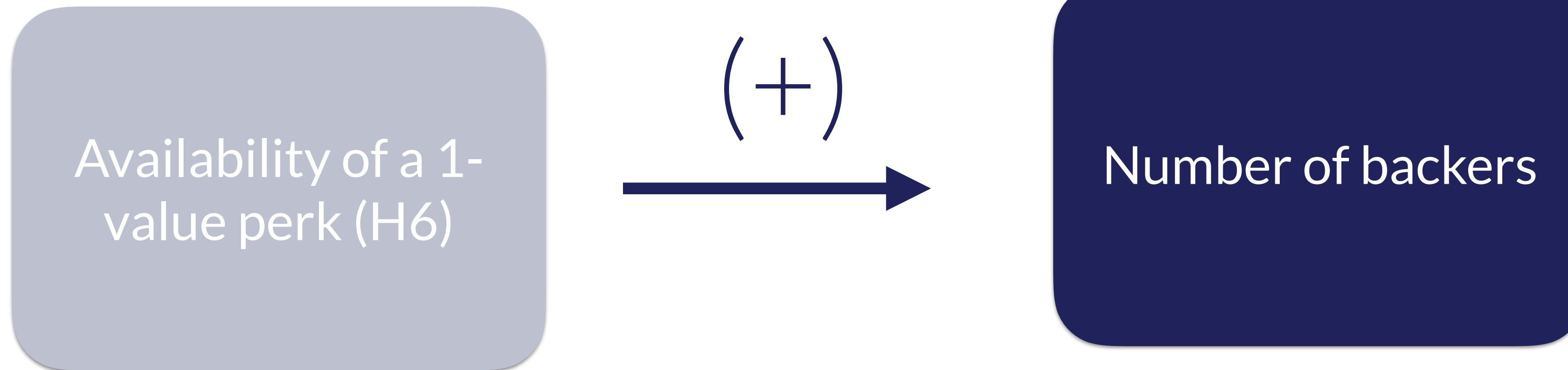
\*Only campaigns of which the `totalNumberBackers` is between  $[Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR] = [0, 174.5]$  have been taken into account.



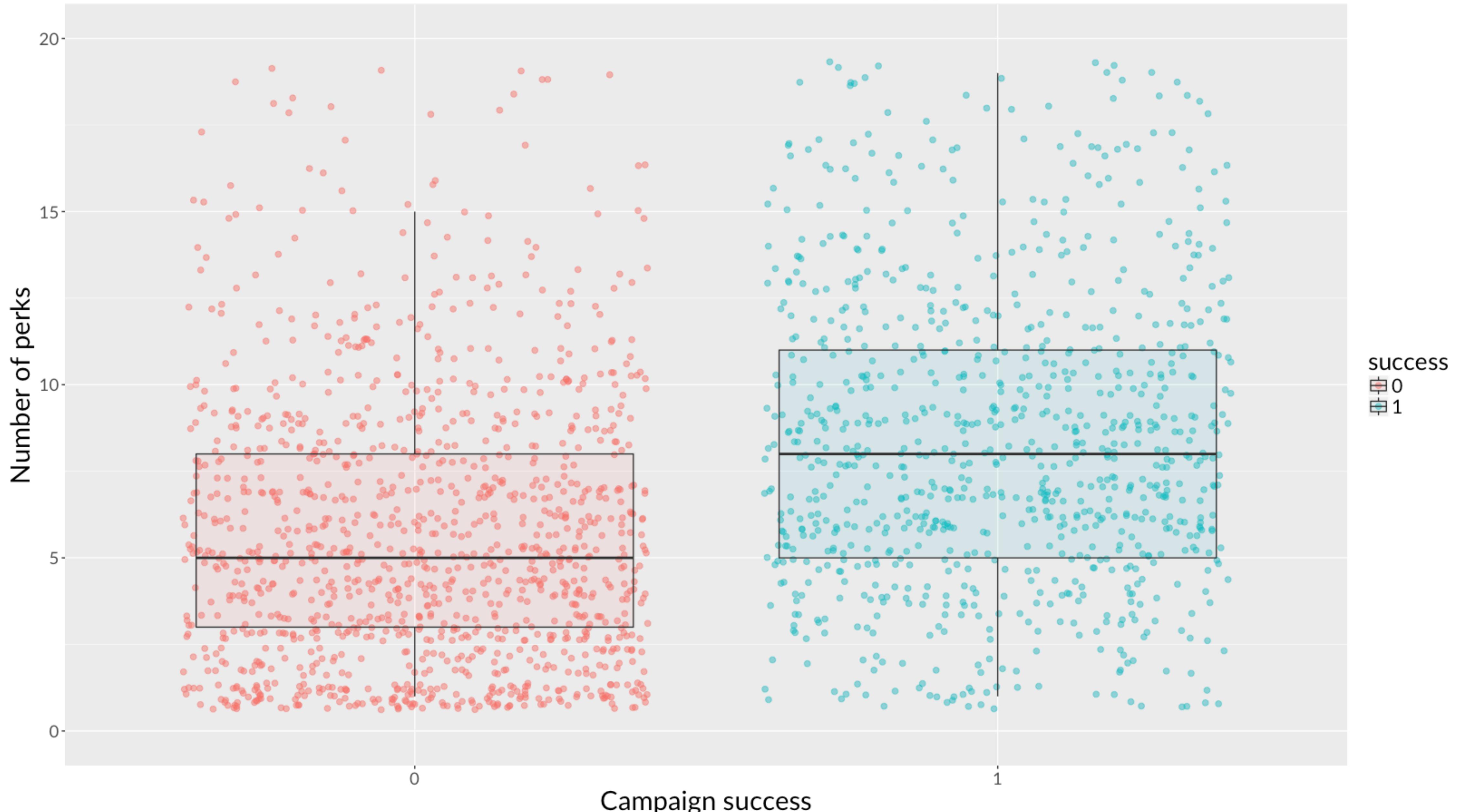
# Significant difference?

**Yes**, p-value: **2.22e-06 (\*\*\*)** - Analysis of Variance (ANOVA)

**Conclusion:** Campaigns with a 1-value perk have significantly more backers.



## Number of perks for unsuccessful (0) and successful campaigns (1)



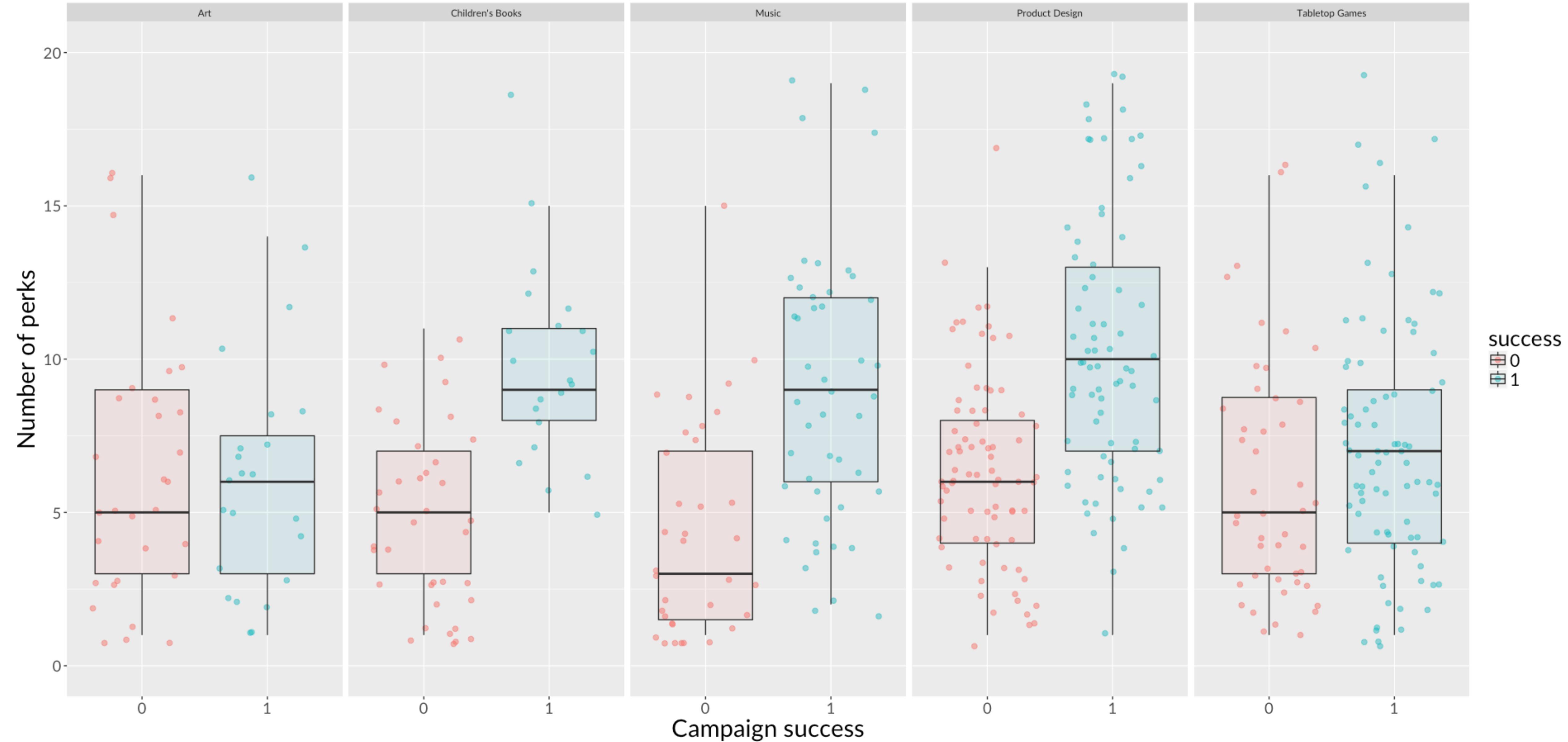


# Statistics number of perks

	Mean*	Median*	Mode*
Successful campaign	8.44	8	6
Unsuccessful campaign	5.67	5	1

\*Only campaigns of which the number of perks is within  $[Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR] = [0, 19]$  have been taken into account.

# Number of perks for unsuccessful (0) and successful campaigns (1) for top-5 categories

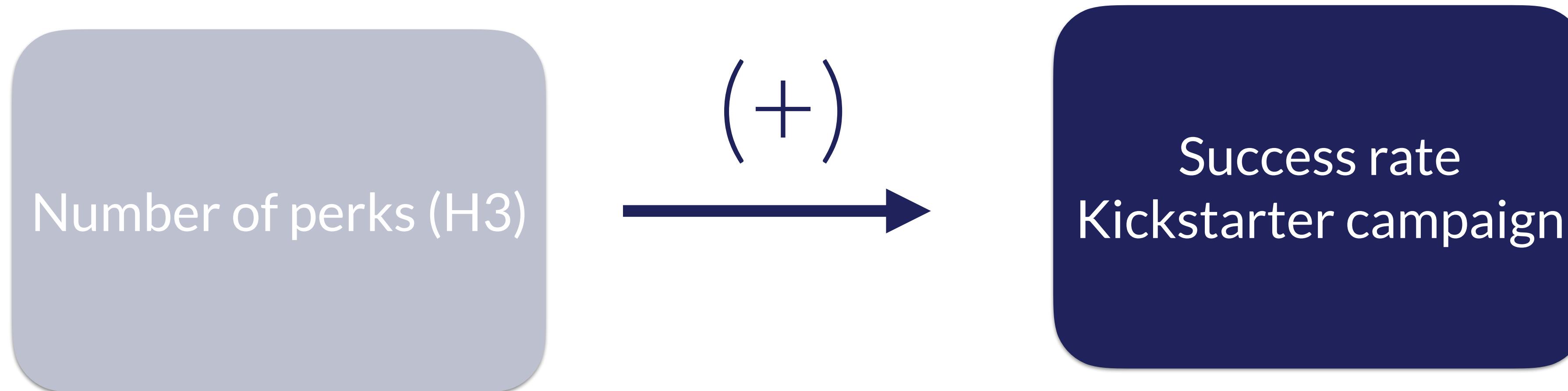




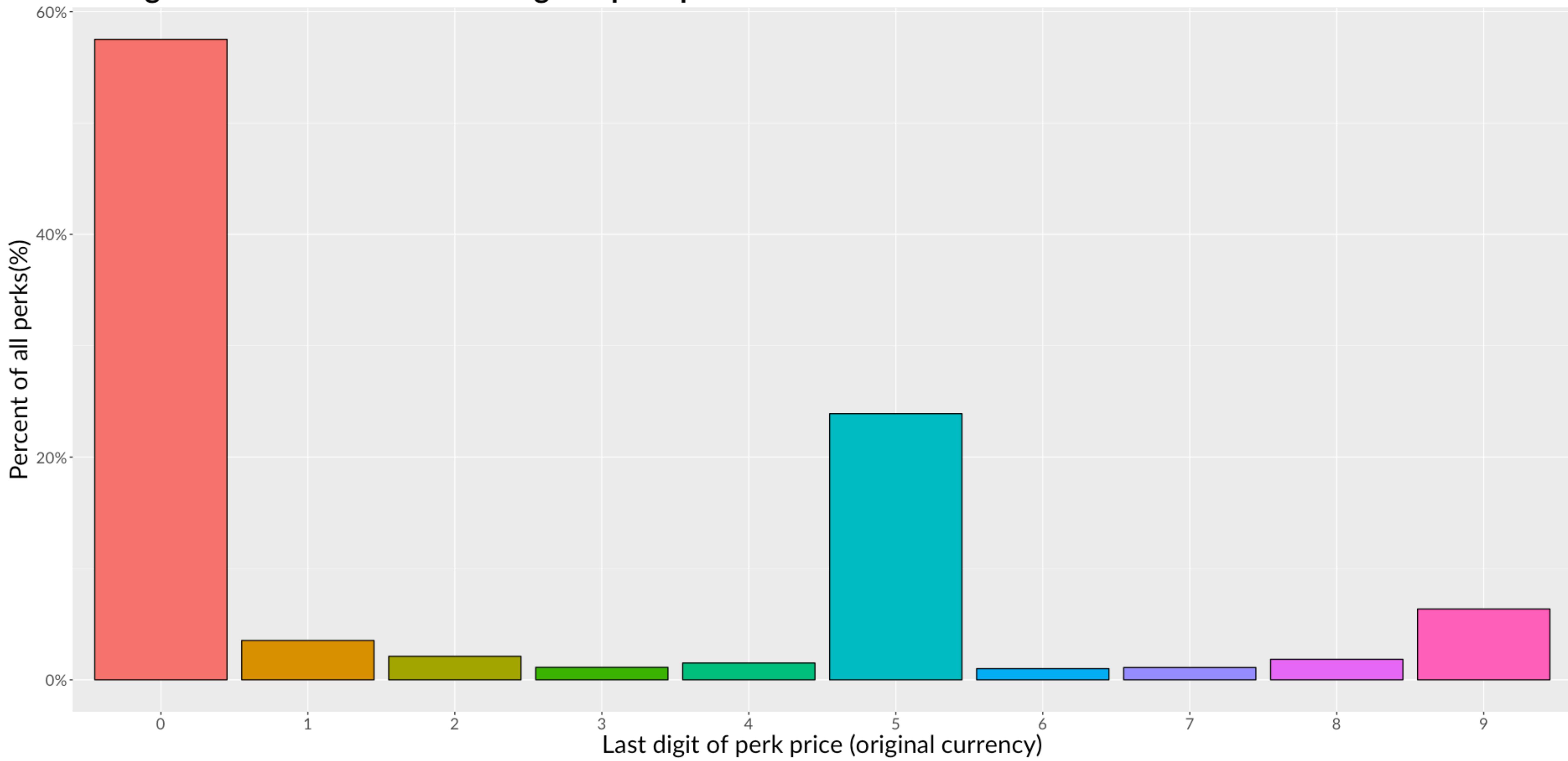
# Significant difference?

**Yes**, p-value: **<2e-16 (\*\*\*)** - Analysis of Variance (ANOVA)

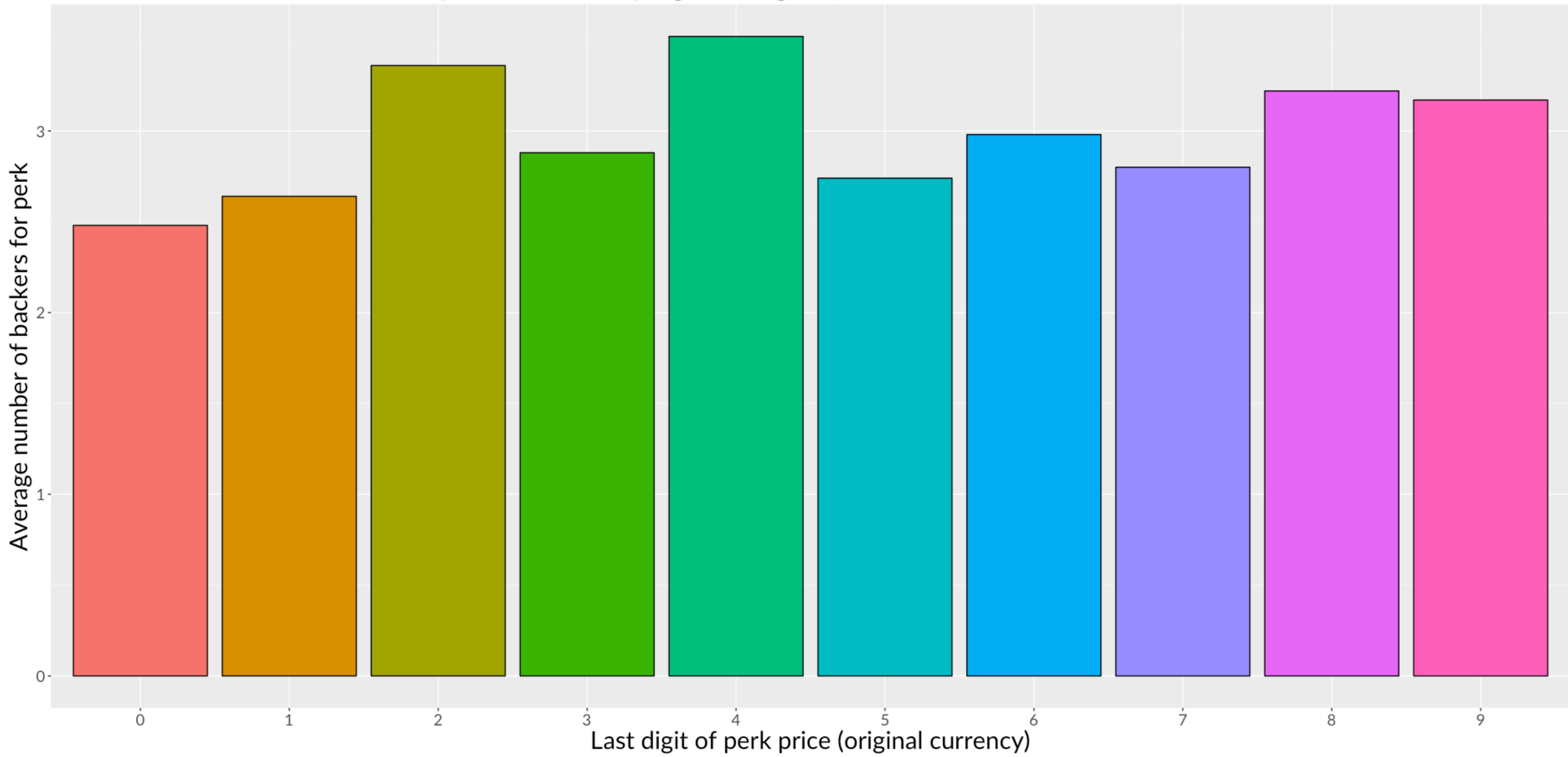
**Conclusion:** Successful campaigns offer significantly more perks than unsuccessful campaigns.



## Histogram of distribution of last digit of perk price



## Mean number of backers for perks with varying last digits (dollar value <500)



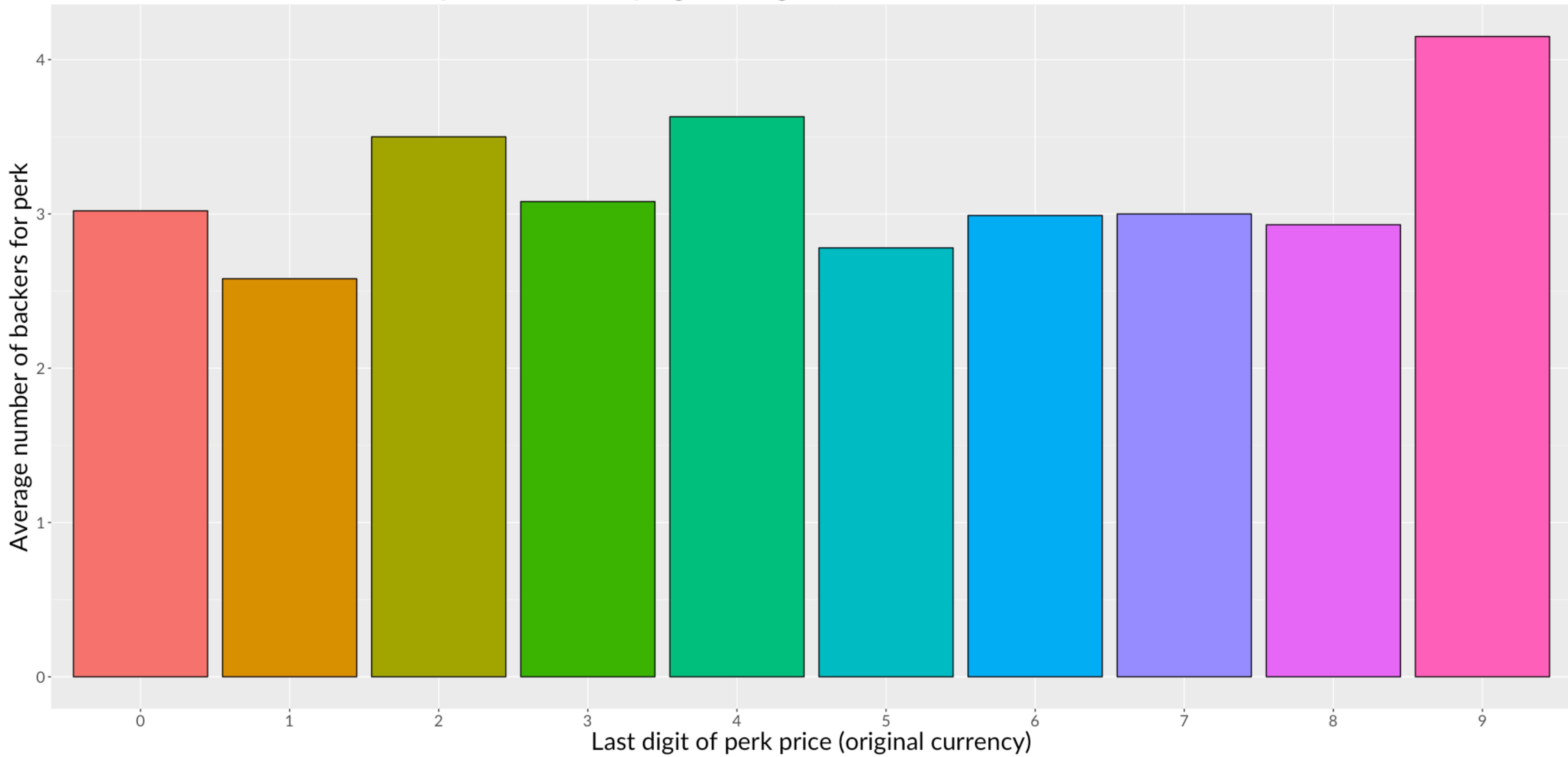


# Different last digit / dollar value

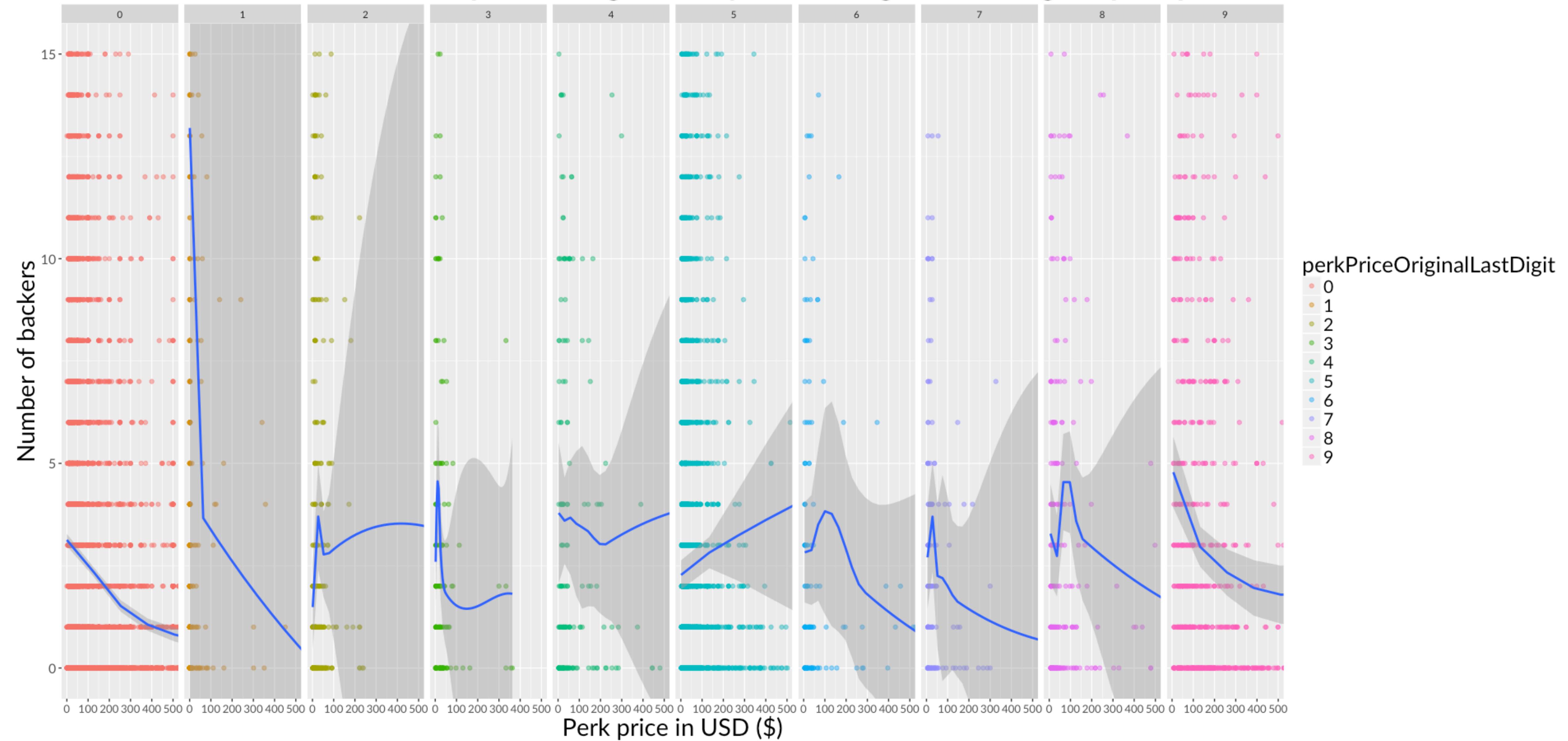
perkPriceOriginal	perkPriceOriginal LastDigit	perkPriceDollar	perkPriceDollar LastDigit
DKK360	0	~ \$ 51	1
\$59	9	\$59	9

perkPriceOriginal	perkPriceOriginal LastDigit	perkPriceDollar	perkPriceDollar LastDigit
DKK361	1	~ \$ 51	1
\$1	1	\$1	1

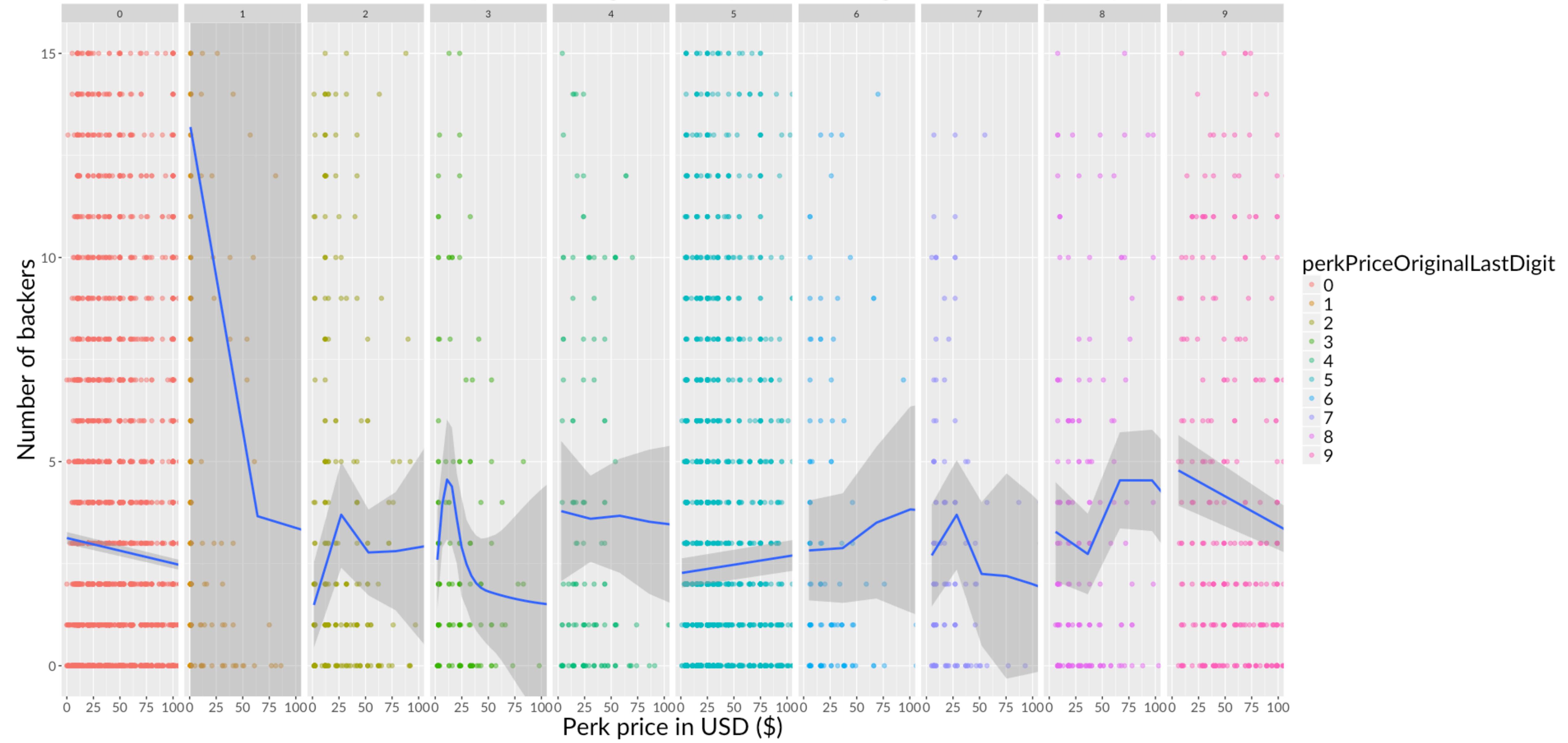
## Mean number of backers for perks with varying last digits (dollar value <50)



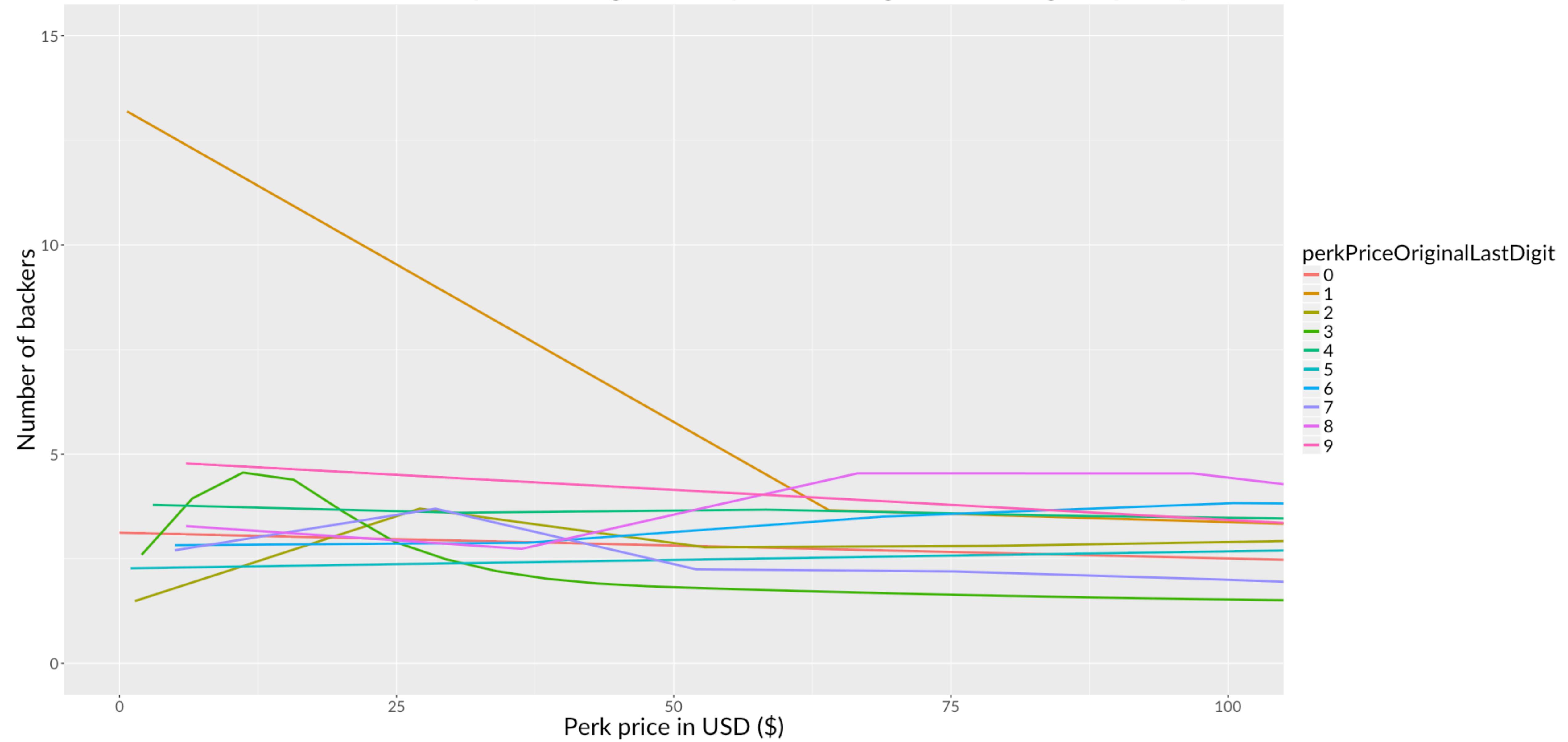
# Number of backers for 0-500\$ perks categorised by the last digit of the original perk price



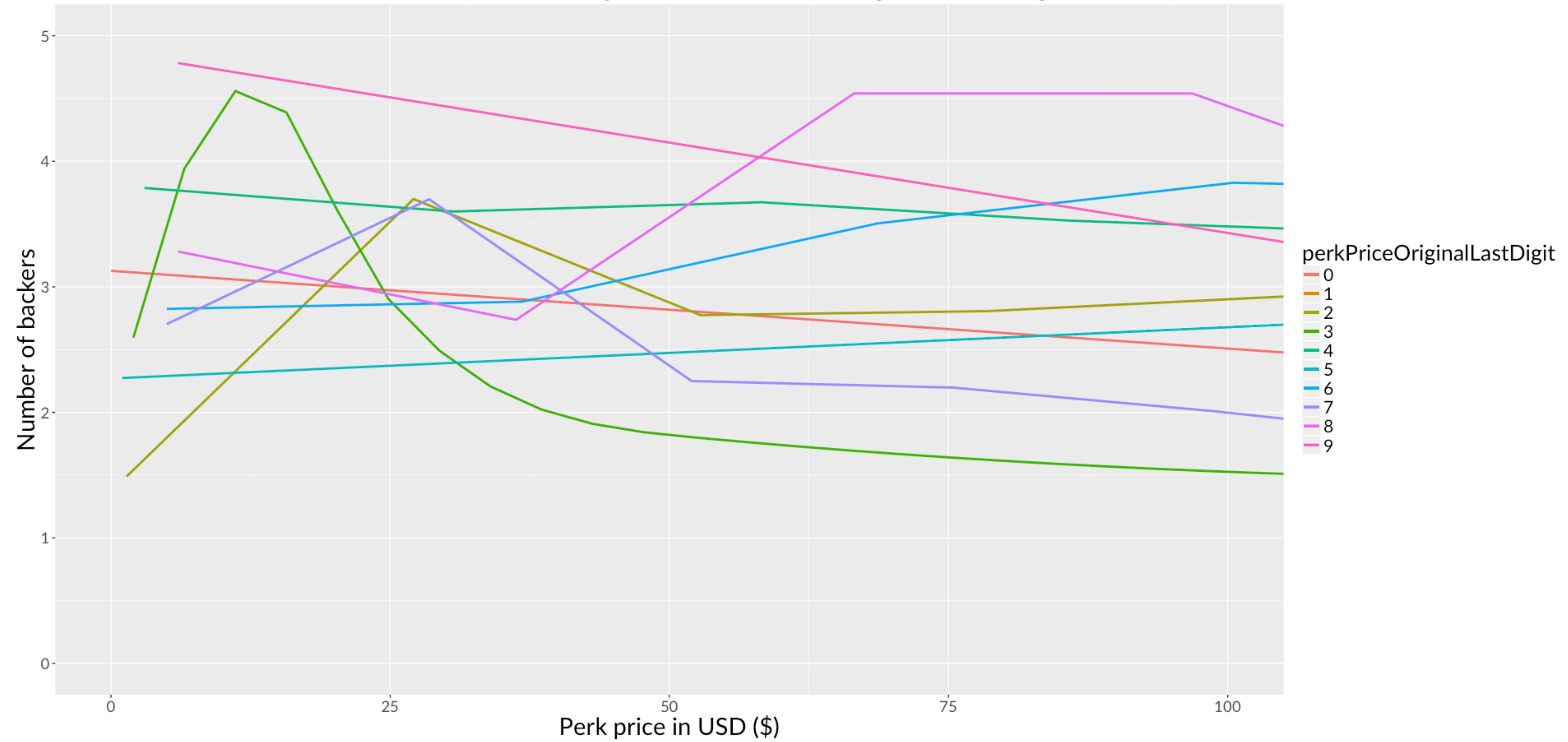
## Number of backers for 0-100\$ perks categorised by the last digit of the original perk price

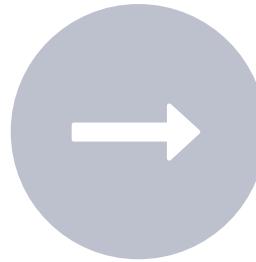


# Number of backers for 0-100\$ perks categorised by the last digit of the original perk price



# Number of backers for 0-100\$ perks categorised by the last digit of the original perk price





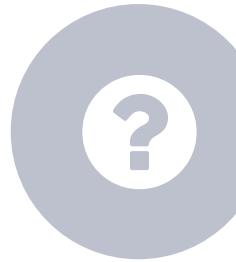
# Suggested control variables

#	Hypothesis
1	Number of campaigns created before by the KS-creator - 1 (rate for success rate of previous campaigns)
2	Price range - 3
3	Campaign category
4	“Project we love” (featured - dummy)
5	Country
6	Number of updates
7	Number of comments
8	Campaign duration
9	Text (# words) - 2
10	Availability of a video (dummy)
11	Number of images



# Current project state

#	Hypothesis	Confirmed	Done	In progress
1	Price level of lowest pledge > (-) > success rate	x	x	
2	Including <= \$25 and >= \$100 perks > (+) > success rate	x	x	
3	Number of perks > (+) > success rate	x	x	
4	Mean pledge > (-) > success rate			x
5	Including 1-value perk > (+) > total number of backers	x	x	
6	Including 1-value perk > (+) > success rate	x	x	
7	Number of added perks after launch > (+) > success rate			x
8	Perk with quantity > (-) > average time period from launch to pledging			x
9	Including perk with quantity limit > (+) success rate			x
10	9 price ending > (+) > average number of pledgers for perk	?	x	
11	Perk with unique price point > (+) > average number of pledgers for perk			x



# Some questions/remarks from my side

---

- Are there relevant hypotheses / research questions **missing**?
- Is it **wise** to **increase the sample size**; to collect more data, basically ignoring Kickstarter's terms' of use? If so, how much additional URLs should be scraped?
- **What types of statistical techniques** should be applied when taking into account the moderators/controls?
- Should the **data collection and data preprocessing** be explained in detail in the **final report** (or appendix) or not?
- Any better way to **formulate “1-value” perks**? (i.e. €1,00; \$1.00; MX\$ 1 etc.)
- How to **combine** the results for all hypotheses **together**? (Mail Madis)
- **Word choice:**  
“Design of the perk system” (company perspective) vs “Perk Pricing Strategy” (creator perspective) - (Mail Madis)



# Some questions/remarks from your side

---

# Thank you!

On to the next part of my data journey...

If any questions/remarks arise, please let me know!

Roy Klaasse Bos

r.j.klaasse.bos@student.tue.nl



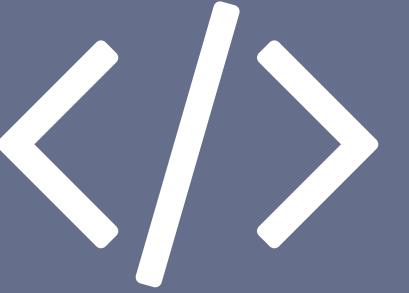
# Table of Contents

---

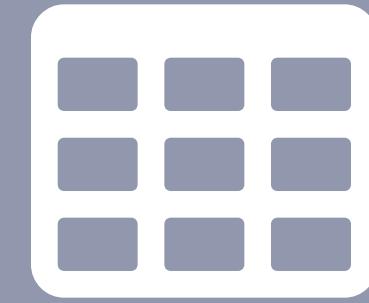
1. Data Collection



2. Data  
preprocessing



3. Aggregated  
statistics



4. Visualisation



- Concrete Kickstart voorbeelden geven; helpt
- Intermediate lag de focus
- Most impressive stuff
- Is Bob bekend met Kickstarter
- De fout die ik eerder maakte: te veel hypotheses (kijken welke data heb ik voorhanden; waar kan ik wat mee)

- Ik vroeg nog extra hypothesised toevoegen. Alsjeblieft focus op 2-3 goede As far as I can see, the overall question is a suitable umbrella over these ten hypotheses sections. And vice versa – ten H is rather unusual, but they are all very specific so here it is actually appropriate to my mind.
- Regressie-analyse geklooí -
- Controls

# Controls

# Table of Contents

Concerning the presentation, I would make it rather research-driven. That means, it largely adopts the structure of the thesis itself, focusing mostly on the rationale for each hypothesis, the method and finally the results. I suggest for that approach because the audience is perfectly okay capturing that stuff very quickly and I think a managerially oriented structure would not be thorough enough.

I would suggest focusing in that presentation on the impressive elements in your work because the second assessor Bob doesn't know much about the thesis yet.

in principle one can discuss forever what to include, how exactly to make the models, what influences what theoretically etc