



# Understanding Google Play Store Application Reviews

**RTime2Shine: Sara Lemus, Zoe Gezelter,  
Sanjay Satish, Lukengu Tshiteya**

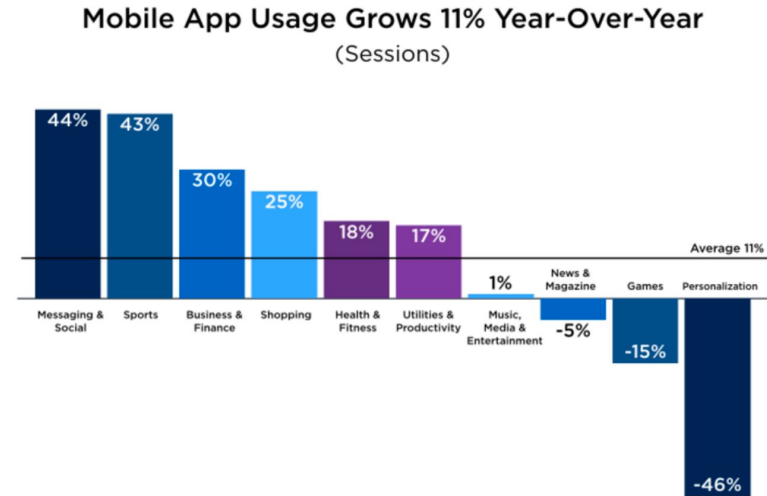
A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting different levels of connectivity or importance. The lines are thin and gray, creating a mesh-like structure.

1.

# Introduction

## Motivation/Goal

- ◎ Understand the significant factors that affect an app's rating
- ◎ Create model that accurately predicts an app's rating given predictors



Source: Flurry Analytics, 2015–2016 Year-Over-Year Session Growth

## Question & Hypothesis

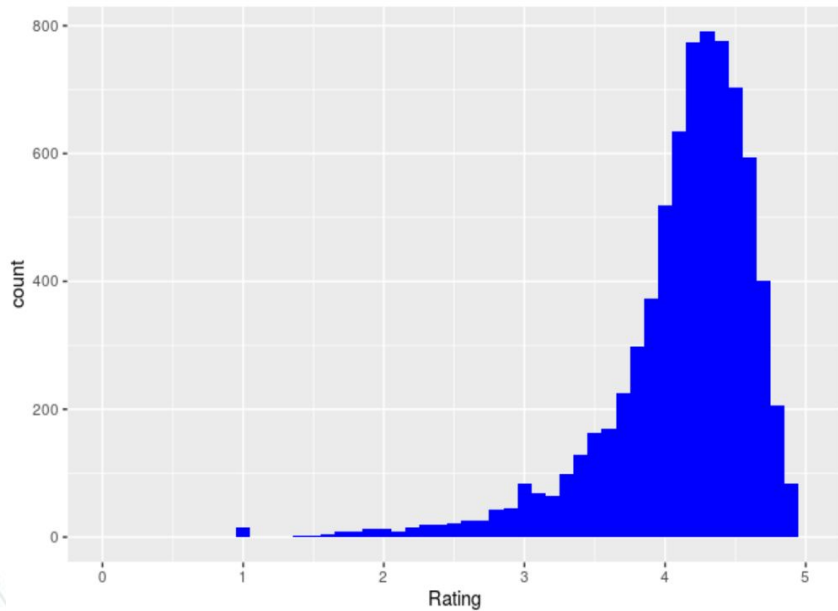
- ◎ What are the relevant factors that affect the rating given for apps in the Google Play store?
- ◎ Hypothesis: Category, Price, Installs, and Content Rating are the predictor variables that will most affect app rating

# Dataset

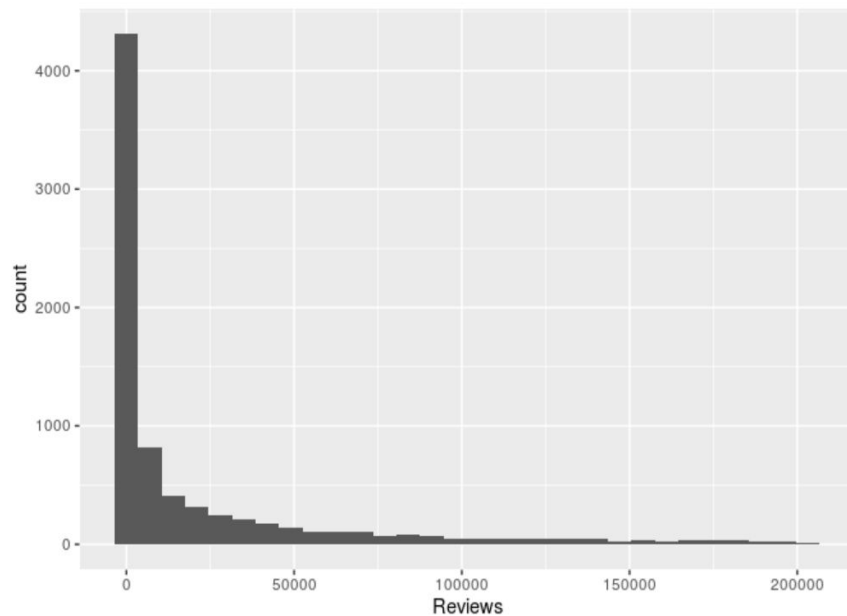
App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
Photo Editor & ...	ART_AND_DESI...	4.1	159	19	10,000	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
Coloring book ...	ART_AND_DESI...	3.9	967	14	500,000	Free	0	Everyone	Art & Design;Pr...	January 15, 2018	2.0.0	4.0.3 and up
U Launcher Lite ...	ART_AND_DESI...	4.7	87510	8.7	5,000,000	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
Sketch - Draw ...	ART_AND_DESI...	4.5	215644	25	50,000,000	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
Pixel Draw - Nu...	ART_AND_DESI...	4.3	967	2.8	100,000	Free	0	Everyone	Art & Design;Cr...	June 20, 2018	1.1	4.4 and up
Paper flowers in...	ART_AND_DESI...	4.4	167	5.6	50,000	Free	0	Everyone	Art & Design	March 26, 2017	1.0	2.3 and up
Smoke Effect P...	ART_AND_DESI...	3.8	178	19	50,000	Free	0	Everyone	Art & Design	April 26, 2018	1.1	4.0.3 and up
Infinite Painter	ART_AND_DESI...	4.1	36815	29	1,000,000	Free	0	Everyone	Art & Design	June 14, 2018	6.1.61.1	4.2 and up
Garden Colorin...	ART_AND_DESI...	4.4	13791	33	1,000,000	Free	0	Everyone	Art & Design	September 20, ...	2.9.2	3.0 and up
Kids Paint Free ...	ART_AND_DESI...	4.7	121	3.1	10,000	Free	0	Everyone	Art & Design;Cr...	July 3, 2018	2.8	4.0.3 and up
Text on Photo - ...	ART_AND_DESI...	4.4	13880	28	1,000,000	Free	0	Everyone	Art & Design	October 27, 2017	1.0.4	4.1 and up
Name Art Phot...	ART_AND_DESI...	4.4	8788	12	1,000,000	Free	0	Everyone	Art & Design	July 31, 2018	1.0.15	4.0 and up
Tattoo Name O...	ART_AND_DESI...	4.2	44829	20	10,000,000	Free	0	Teen	Art & Design	April 2, 2018	3.8	4.1 and up
Mandala Colori...	ART_AND_DESI...	4.6	4326	21	100,000	Free	0	Everyone	Art & Design	June 26, 2018	1.0.4	4.4 and up
3D Color Pixel b...	ART_AND_DESI...	4.4	1518	37	100,000	Free	0	Everyone	Art & Design	August 3, 2018	1.2.3	2.3 and up
Learn To Draw K...	ART_AND_DESI...	3.2	55	2.7	5,000	Free	0	Everyone	Art & Design	June 6, 2018	NaN	4.2 and up
Photo Designer ...	ART_AND_DESI...	4.7	3632	5.5	500,000	Free	0	Everyone	Art & Design	July 31, 2018	3.1	4.1 and up
350 Diy Room ...	ART_AND_DESI...	4.5	27	17	10,000	Free	0	Everyone	Art & Design	November 7, 20...	1.0	2.3 and up

# EDA

Distribution of App Ratings



Distribution of Reviews



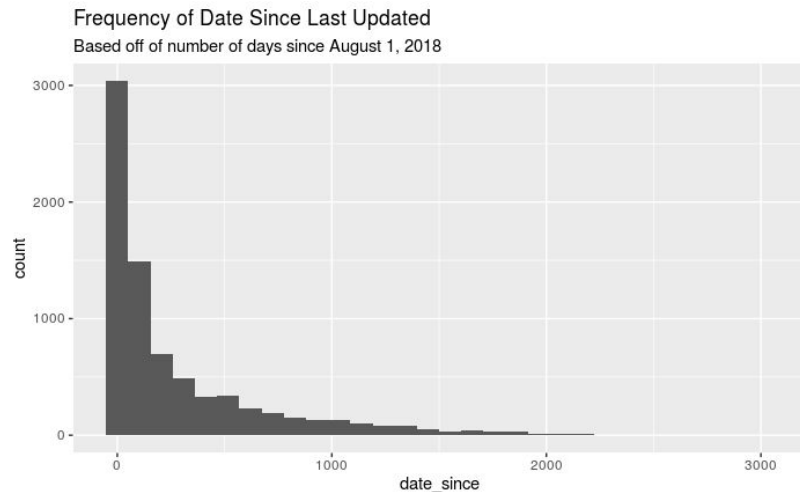
A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting different levels of connectivity or importance. The lines are thin and gray, creating a mesh-like structure.

# 2. **Model**

## Data Wrangling

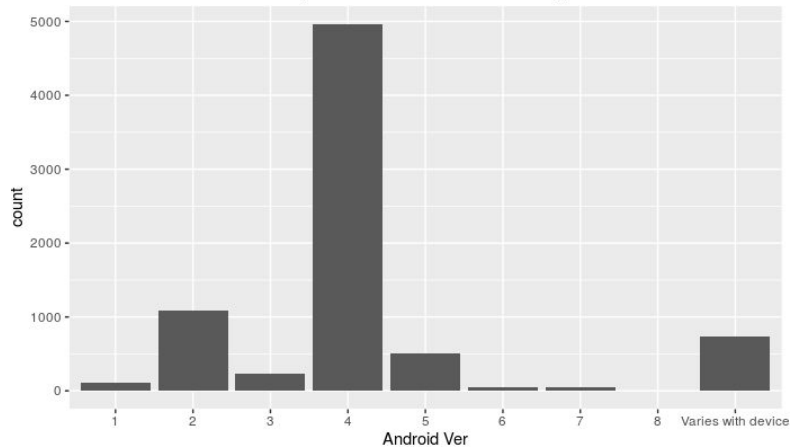
- Removed Type and Genre
- Created days since last update
- Binned additional variables
- Major changes: condensed category, Android Version to keep model as simple as possible and avoid overfitting

date_new	date_since
2018-01-07	213
2018-01-15	205
2018-08-01	7
2018-06-20	49
2017-03-26	500

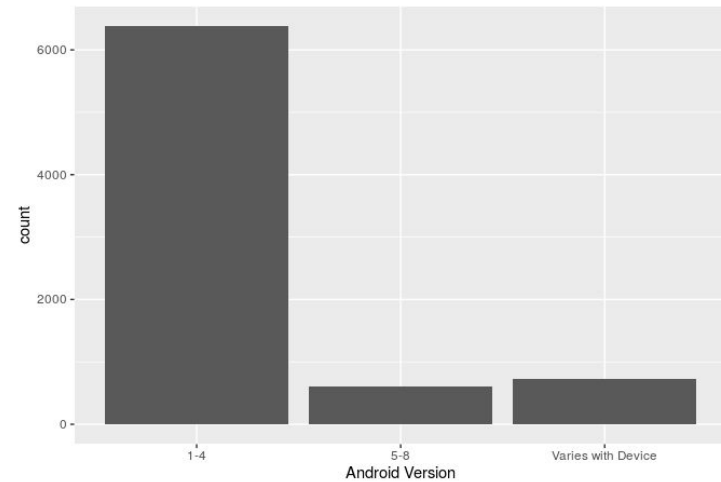




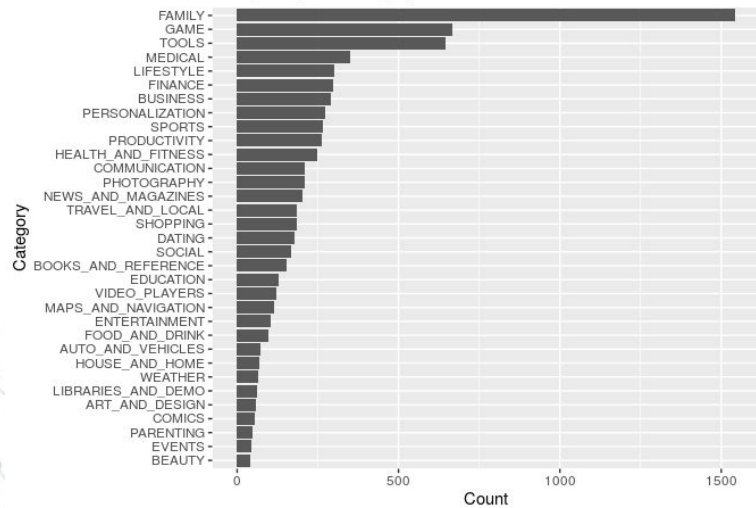
### Distribution of Latest Compatible Android Version of App



### Distribution of Android Version

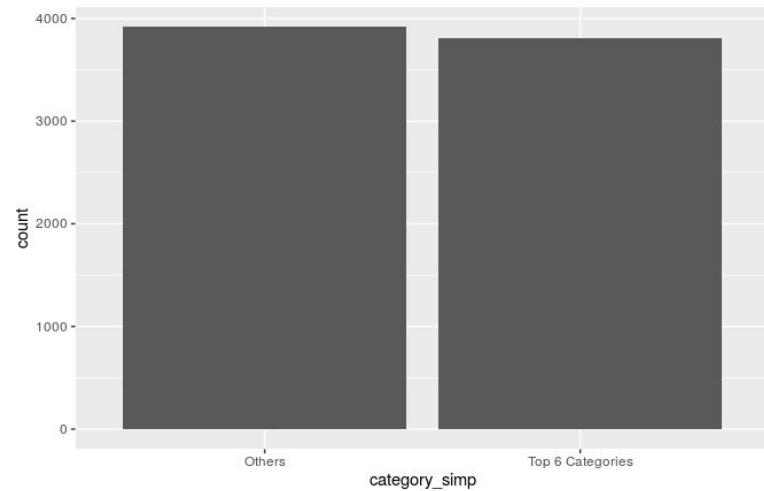


### Frequency of Category



### Frequency of Category

#### Simplified Version



## Multiple Linear Regression

- MLR using AIC backwards selection
- No interaction effects, all p-values less than 0.05

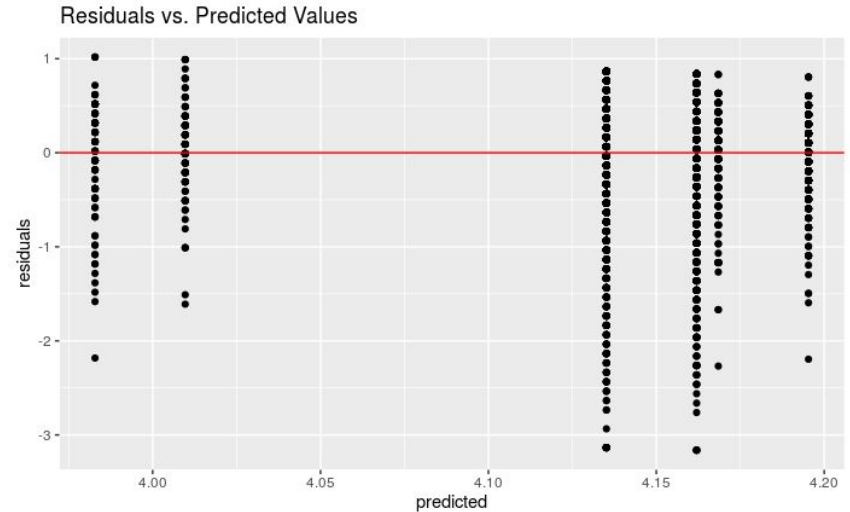
**Final model:**  $4.01 - 0.027(\text{category\_simpTop 6 Categories}) + 0.152(\text{SizeLess than 100 MB}) + 0.186(\text{SizeVaries with device})$

Term	Estimate	Standard Error	Statistic	P Value
Intercept	4.0096938	0.0380781	105.301857	0.0000000
category_simpTop 6 Categories	-0.0268642	0.0126225	-2.128288	0.0333448
SizeLess than 100 MB	0.1523095	0.0381560	3.991756	0.0000662
SizeVaries with device	0.1856302	0.0418267	4.438079	0.0000092

## Model Assumptions

- Compared to other models, AIC conformed the most to assumptions
- No glaring overall issues with violations
- 363 with large residuals, 214 high leverage points, but no Cook's D/ VIF concerns

### Constant Variance

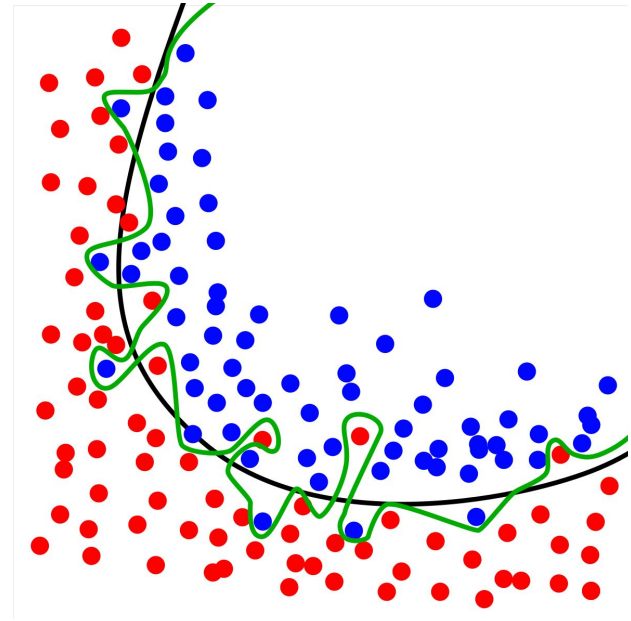


A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting different levels of connectivity or importance. The lines are thin and gray, creating a mesh-like structure.

# 3. **Limitations**

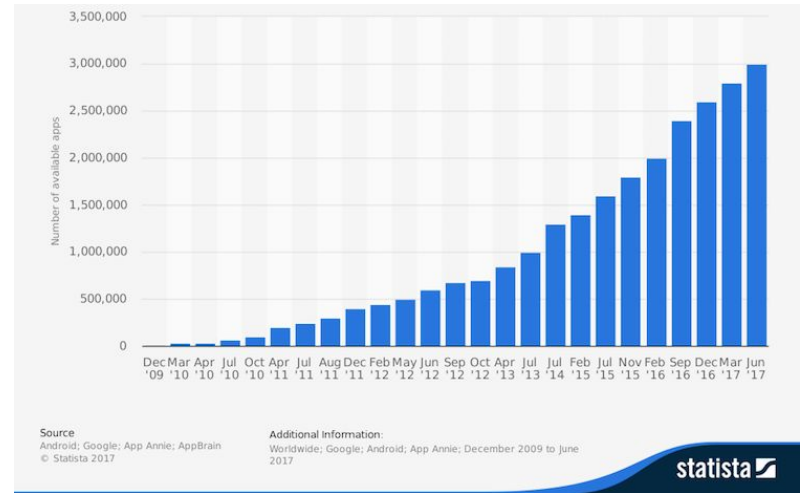
## Overfitting

- Variables such as `categories` have **many levels** and experience **tradeoffs** from releveling
- Model may be "**memorizing**" our data rather than assessing the variability in response through **particular trends** in our predictors
  - Hurts **generalizability**

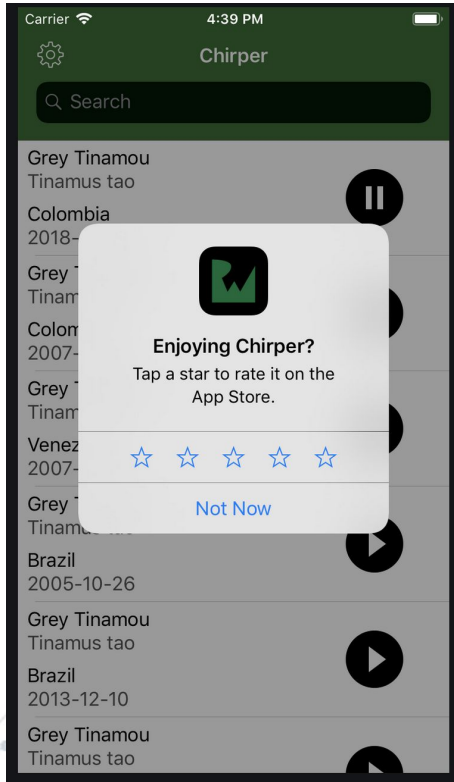


## Dynamism

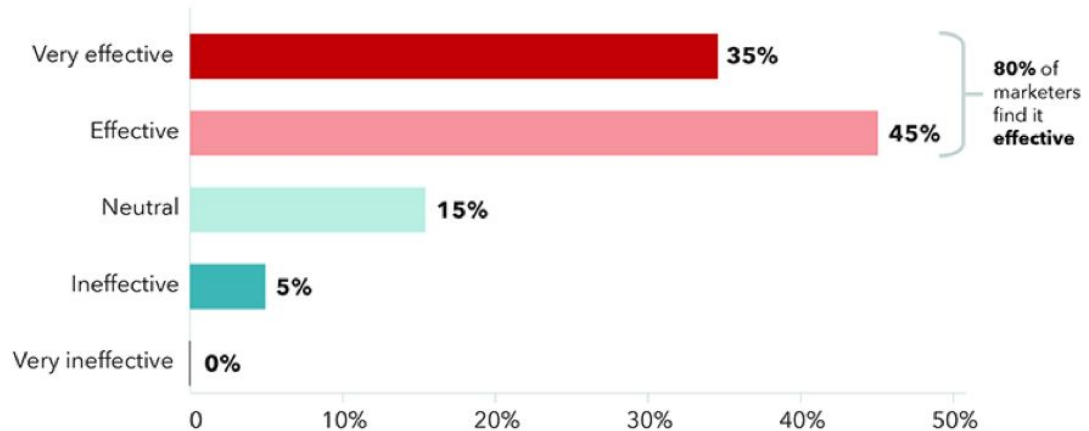
- ◎ The Google Play Store is a **dynamic** environment
  - Categories, diversity in app offerings, etc. **change(s) constantly**
- ◎ A singular model based upon a singular dataset is **not necessarily** the best representation of this if our goal is prediction



# External Factors



## HOW EFFECTIVE IS INFLUENCER MARKETING?



mediakix

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, while others are smaller and solid. The lines are thin and gray, connecting the nodes in a non-linear fashion. The overall shape of the network is roughly triangular, pointing towards the top-left.

4.

# Conclusion and Discussion



## REVIEWS

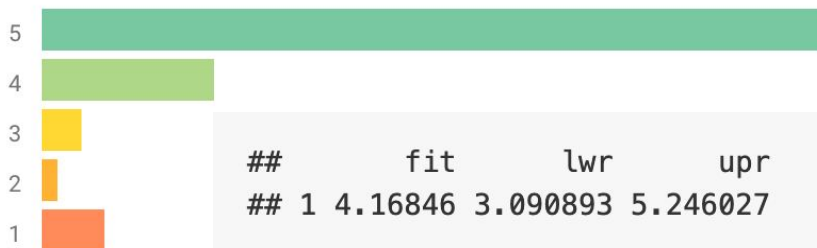
 Review Policy



4.5



225,537 total



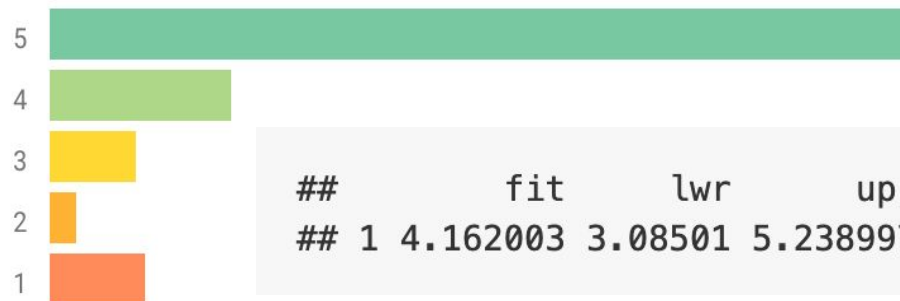
##	fit	lwr	upr
## 1	4.16846	3.090893	5.246027



4.3



43,882 total



##	fit	lwr	upr
## 1	4.162003	3.08501	5.238997

# Predictions

# Facebook buys WhatsApp for \$19 billion

by Adrian Covert @CNNTech

February 19, 2014: 6:54 PM ET

Recommend 39



The New York Times

## *A Taste of Lime: Uber Invests in an Electric Scooter Company*



NEWS | UNIVERSITY

## How the Jailbreak computer game made sophomore Alex Balfanz millions



Alex Balfanz made enough money to pay for four years of tuition at Duke and more with his hit computer game Jailbreak.

Photo by [Simran Prakash](#) | The Chronicle

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting a hierarchical or central structure. The lines are thin and gray, connecting the nodes in a non-linear fashion.


# Questions?

Thank you  
- RTime2Shine

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of nodes connected by lines, with some nodes being larger and more prominent than others, creating a sense of depth and connectivity.

A decorative network diagram in the top-left corner of the slide. It consists of a series of interconnected nodes and edges, forming a complex web-like structure. The nodes are represented by small circles, some of which are shaded in gray, while others are white with a gray outline. The edges are thin gray lines connecting the nodes. The overall shape of the diagram is roughly triangular, pointing towards the top-left corner.

# References

- <https://www.nytimes.com/2018/07/09/technology/uber-invests-lime-electric-scooter.html>
  - <https://money.cnn.com/2014/02/19/technology/social/facebook-whatsapp/>
  - <https://www.dukechronicle.com/article/2018/12/how-the-jailbreak-computer-game-made-sophomore-alex-balfanz-millions>
  - [https://play.google.com/store?hl=en\\_US](https://play.google.com/store?hl=en_US)
  -
- 
- A decorative network diagram in the bottom-right corner of the slide. It is similar in style to the one in the top-left, featuring a web of interconnected nodes and edges. The nodes are small circles, some shaded gray and some white with gray outlines, connected by thin gray lines. The diagram is positioned in the bottom-right corner, mirroring the placement of the other decorative element.