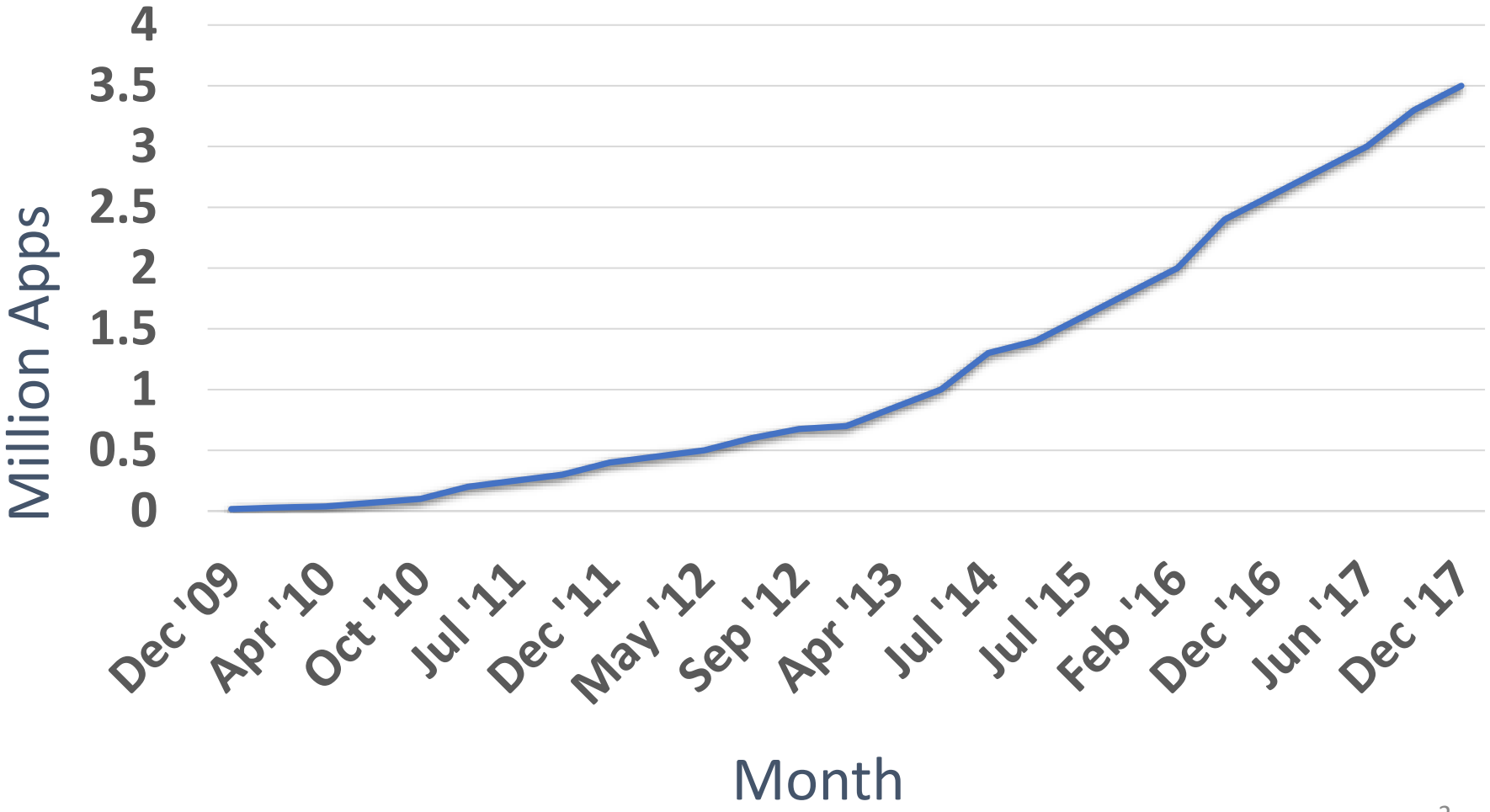
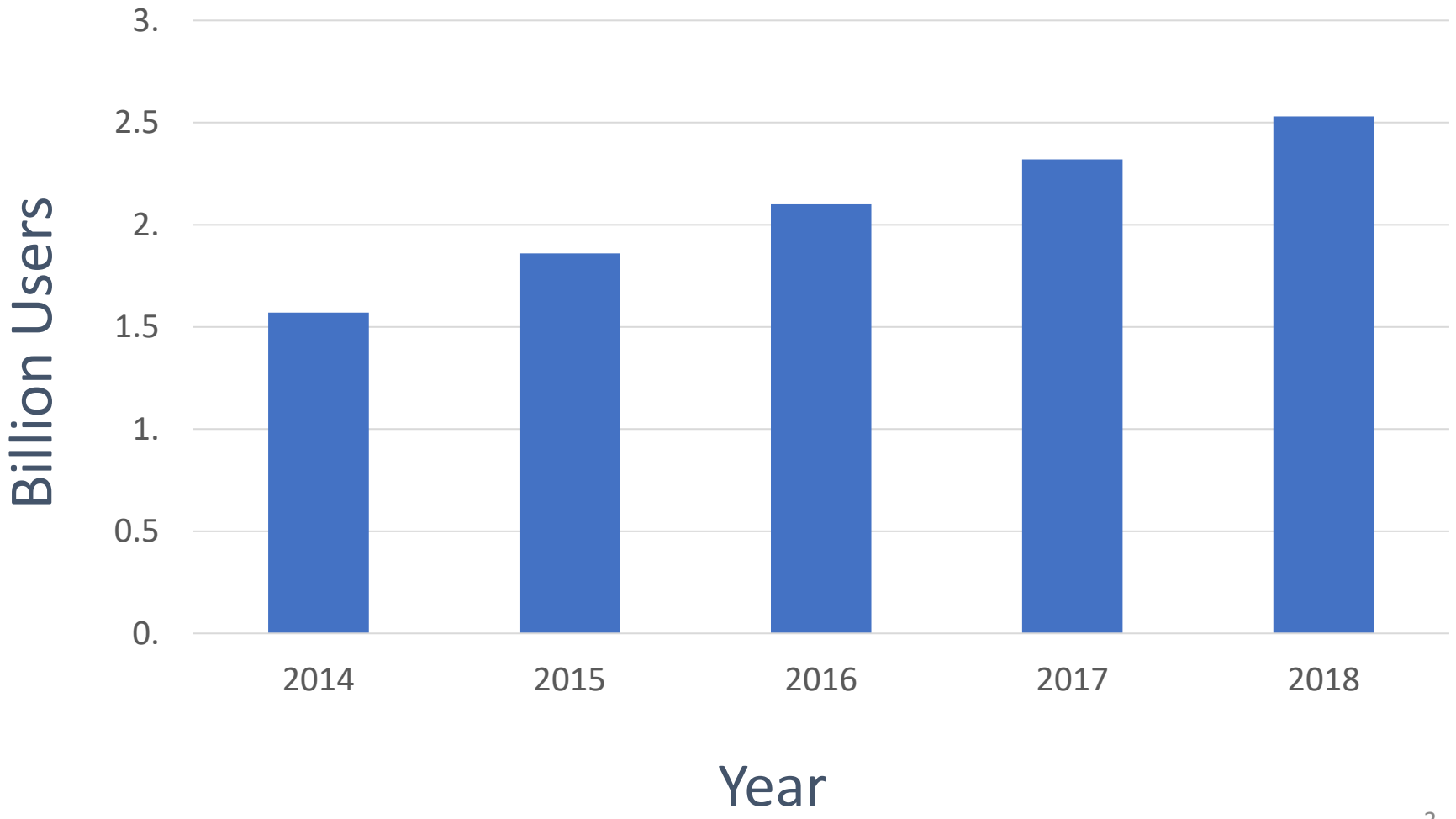
Ehsan Noei
Daniel Alencar da Costa
Ying (Jenny) Zou

# Winning the app production rally

# The number of mobile apps continues to rise



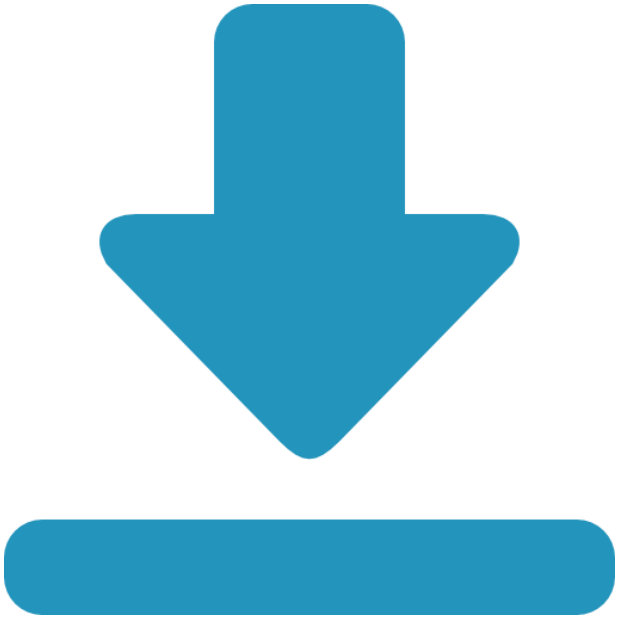
# The numbers of smartphone users is raising



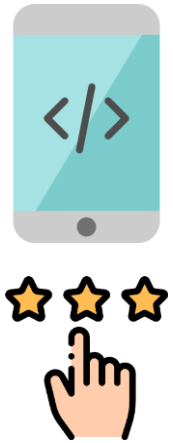
# The app markets are becoming more competitive

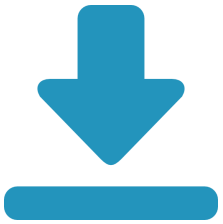
Image Credit: http://pixabay.com/

# Prior studies have investigated the success of mobile apps

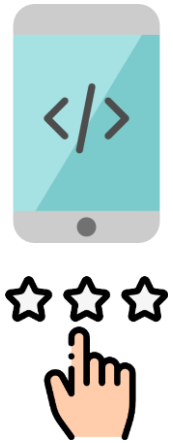


# Studied success factors is not as precise as rankings

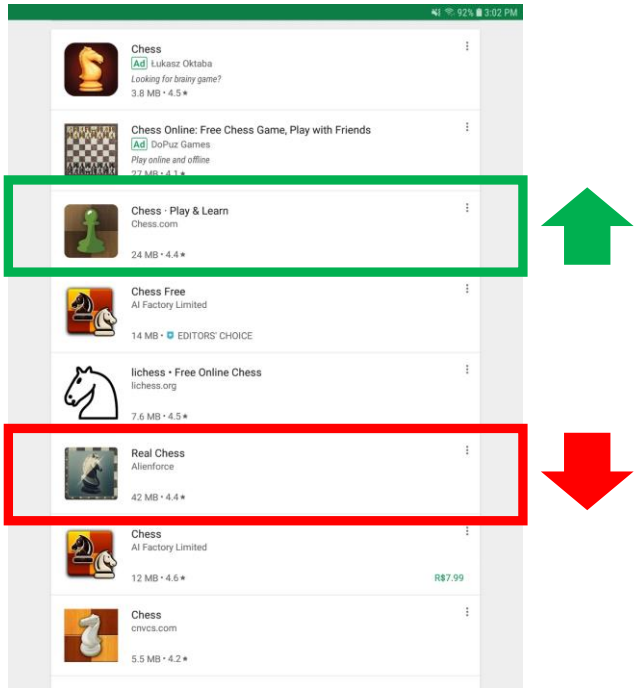




100 Downloads

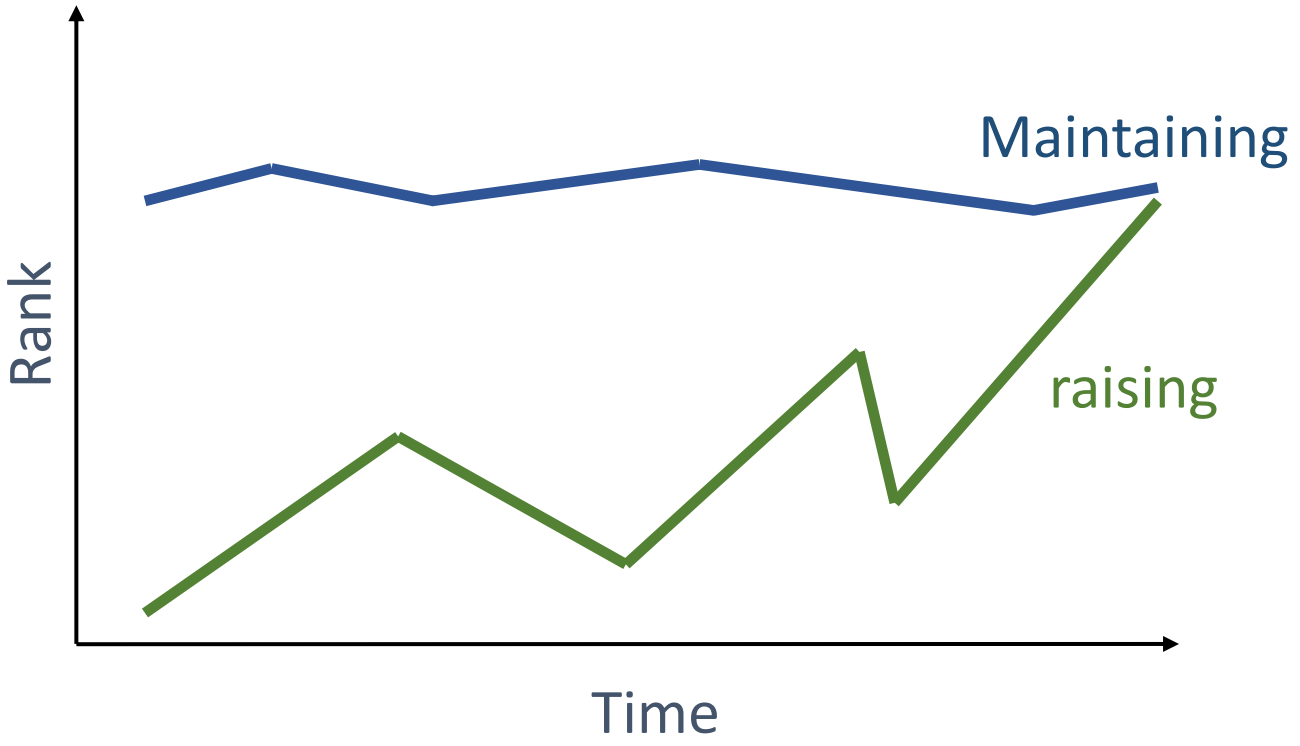


5,000 Downloads



# Rank Trend

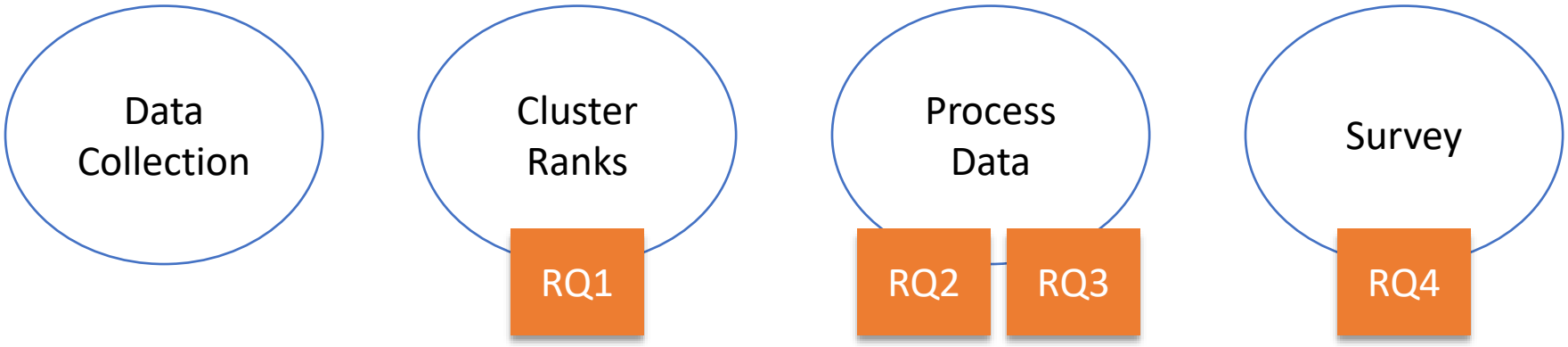
- **Definition.** Rank trend is the evolution of ranks over time.



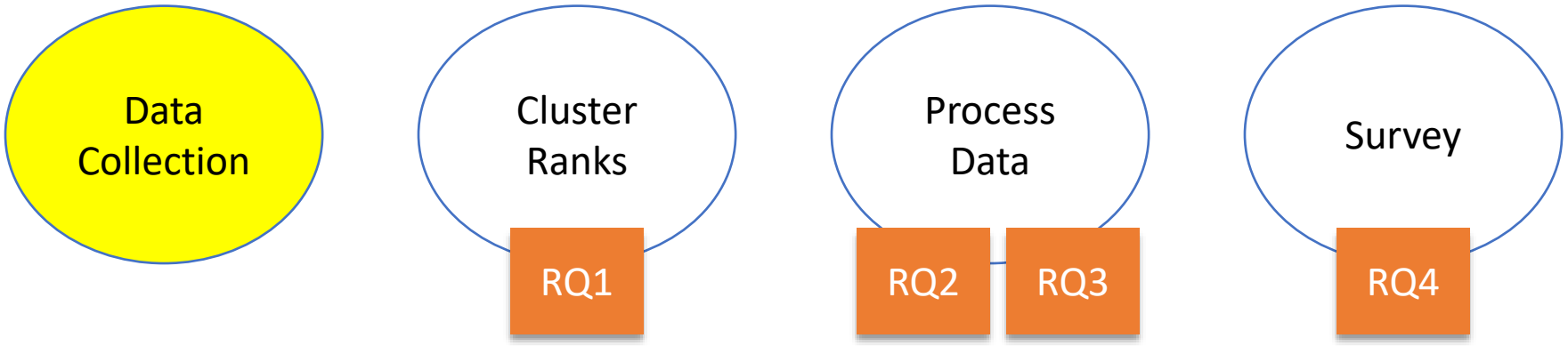
# 900 apps in 30 top areas containing 4,878,011 reviews

| Area | #Versions | #User-reviews | Area | #Versions | #User-reviews |
|---|---|---|---|---|---|
| Airline | 422 | 51,408 | Health | 263 | 88,306 |
| Bible | 339 | 92,075 | Mailbox | 339 | 56,391 |
| Budget | 206 | 18,356 | Messaging | 525 | 230,248 |
| Calculator | 294 | 7,316 | Movie | 279 | 164,195 |
| Calling | 658 | 210,483 | News | 690 | 186,827 |
| Camera | 466 | 55,383 | Paint | 122 | 6,674 |
| Chess | 186 | 22,811 | Piano | 202 | 27,909 |
| Cloud | 546 | 162,591 | Radio | 504 | 238,336 |
| Coupon | 348 | 61,160 | Reminder | 292 | 108,461 |
| Dating | 916 | 165,869 | Sleep | 159 | 42,713 |
| Dictionary | 292 | 74,073 | Spy Phone | 162 | 27,328 |
| Emoji | 101 | 4,105 | Talking Pet | 235 | 31,662 |
| Fitness | 272 | 84,436 | Translator | 206 | 36,375 |
| GPS | 630 | 151,844 | Weather | 845 | 178,968 |
| Grocery List | 264 | 50,944 | Weight Loss | 278 | 58,590 |

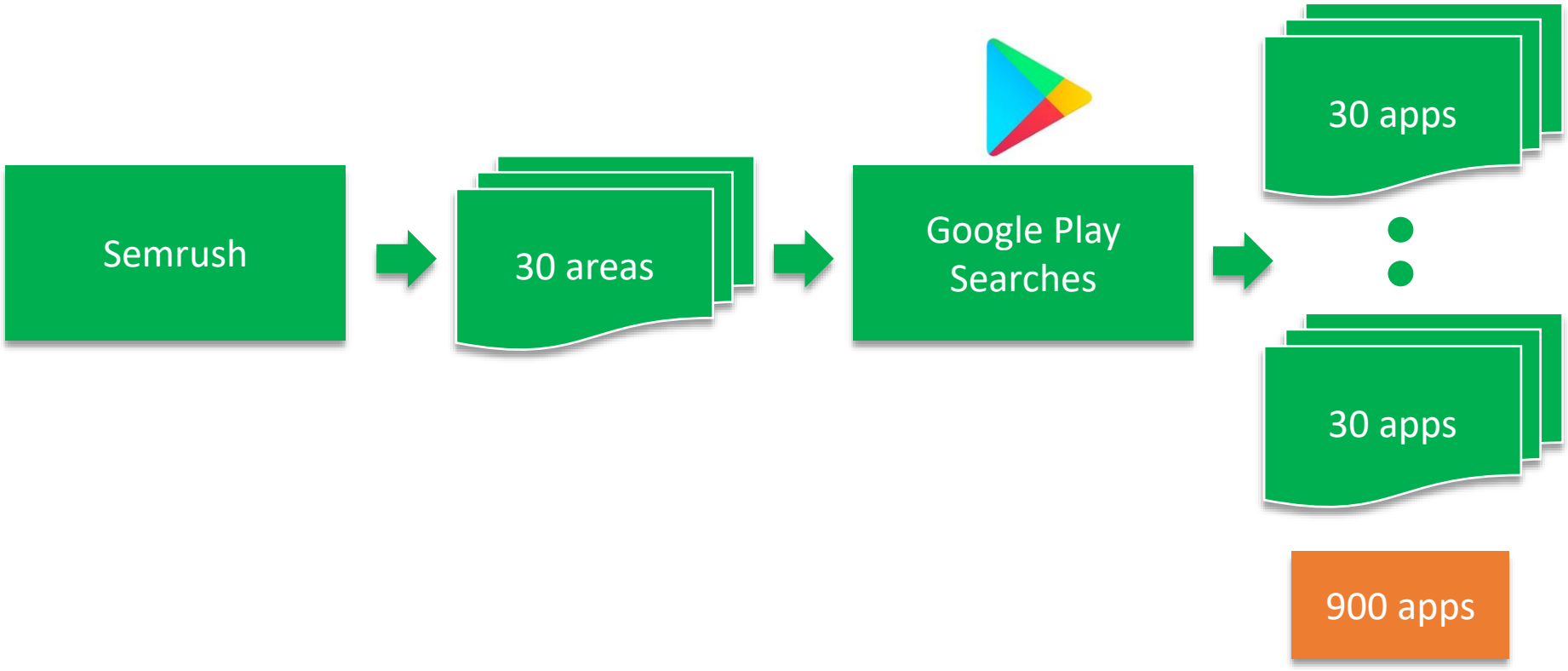
# Study Setup


Data Collection
Cluster Ranks
RQ1
Process Data
RQ2
RQ3
Survey
RQ4

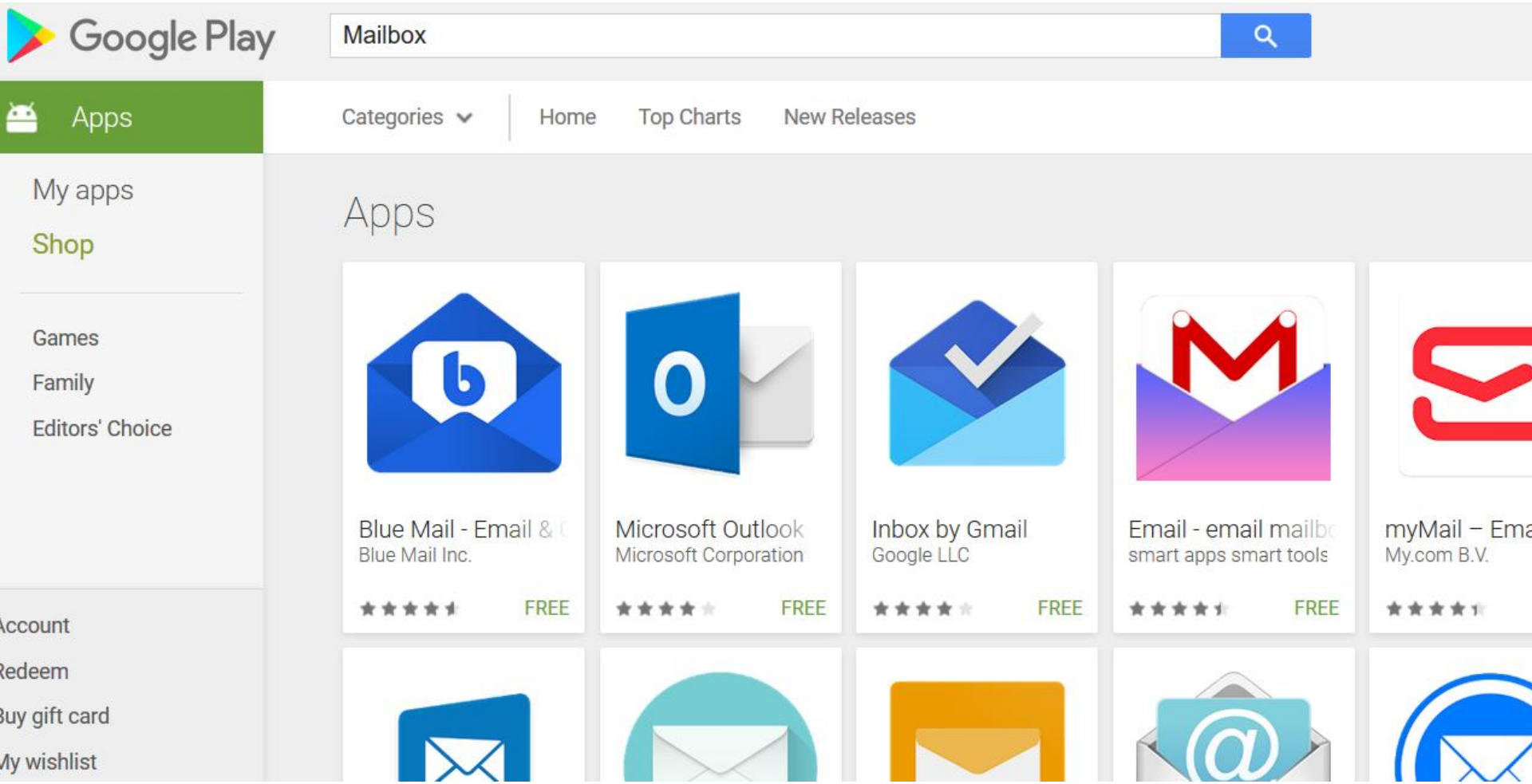
# Data Collection



# Data Collection



# Data Collection



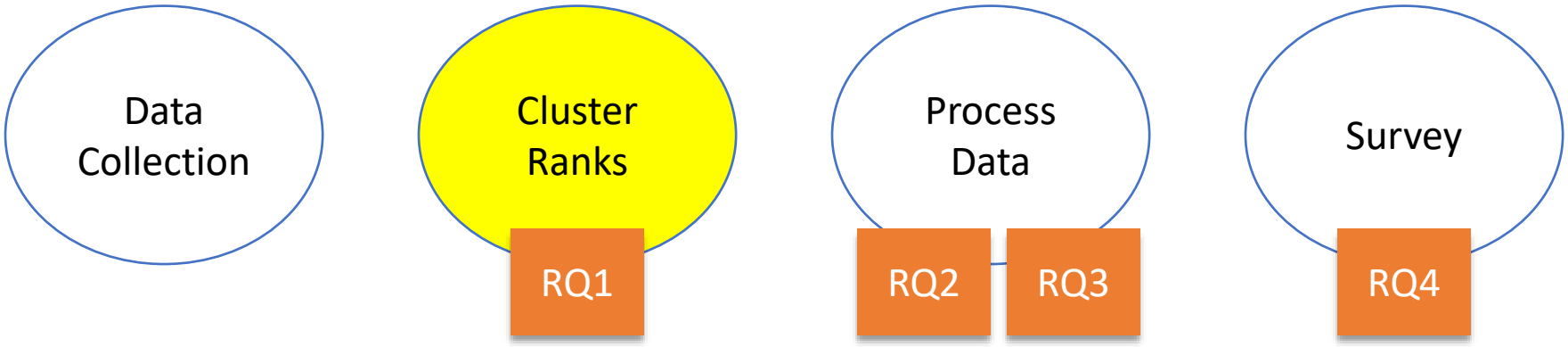
# Data Collection

- Data collection process took over 2 years, so that we would not lose any information.

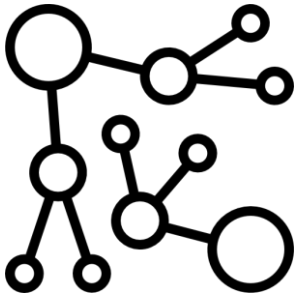
# Data Pre-Processing

- ✓Removing Non-English User-reviews
- ✓Filtering Out Uninformative User-reviews
- ✓Correcting Typos and Informal Vocabularies
- ✓Breaking Down App Descriptions and Release Notes
- ✓Resolving Synonyms
- ✓Resolving Negations
- ✓Topic Modeling

# RQ1: What are the rank trends of mobile apps?

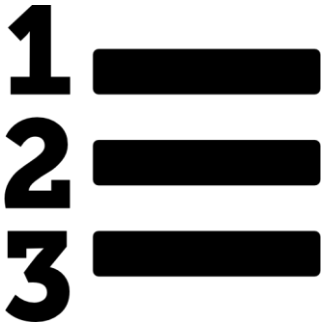


# Time Series Clustering



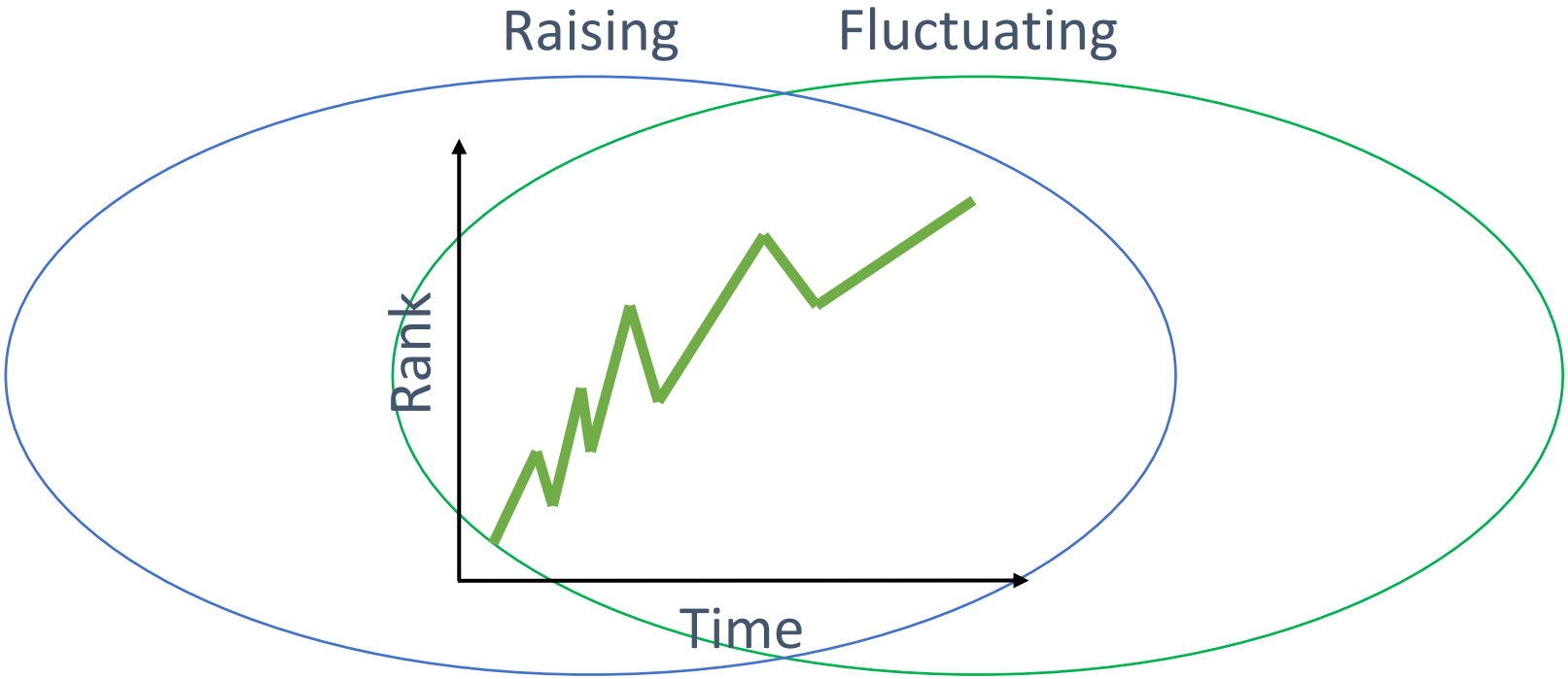
Clustering algorithm

Distance measurement method

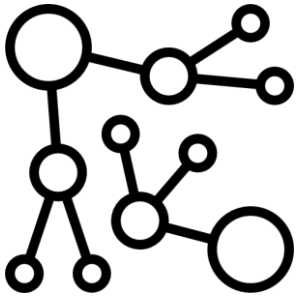



Number of clusters

# Clustering Algorithm: Fuzzy

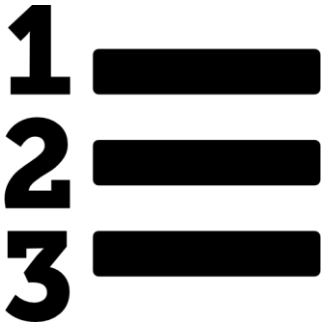


# Time Series Clustering



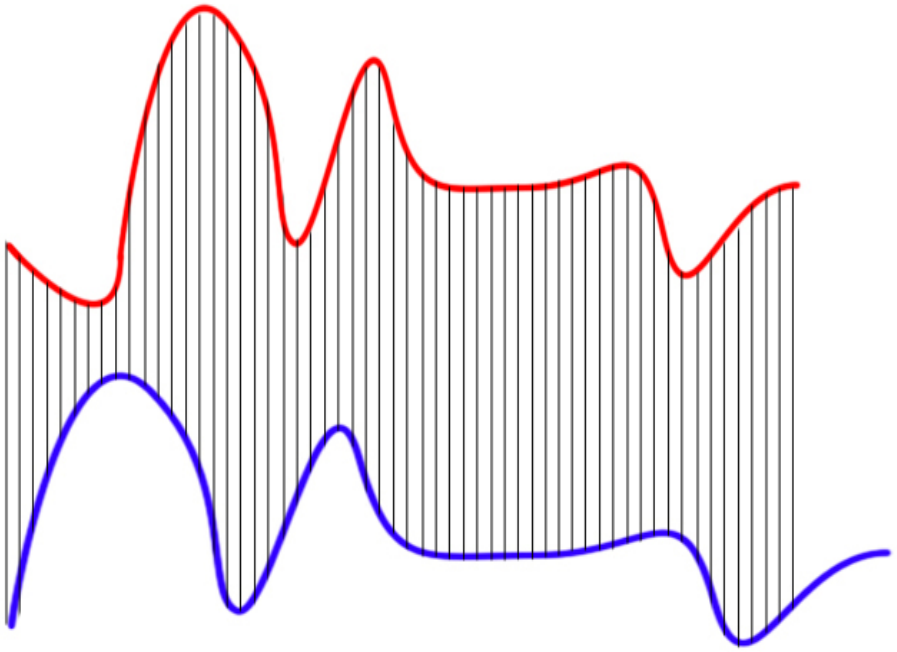
Clustering algorithm

Distance measurement method

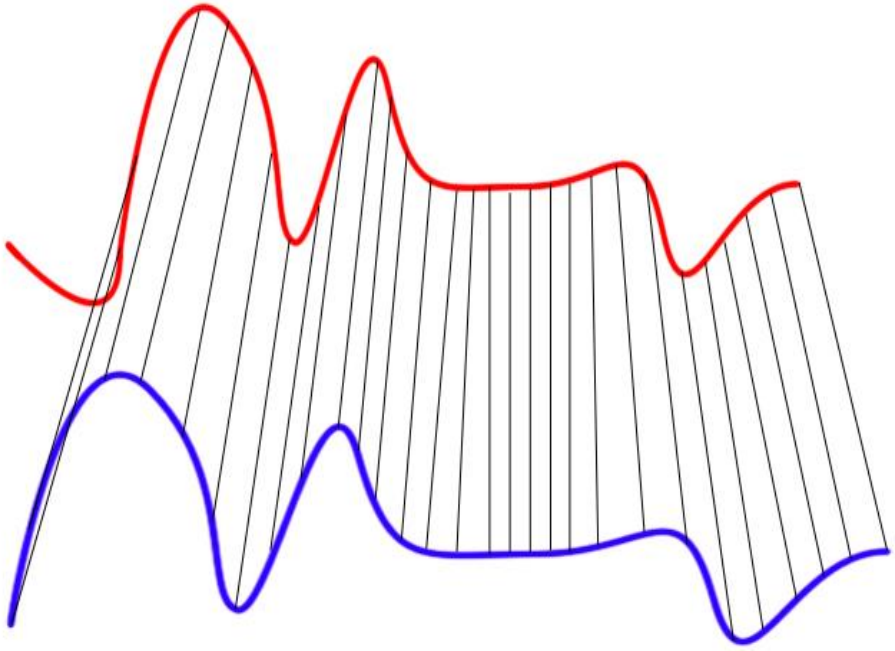



Number of clusters

# Distance Measurement Method:
# Dynamic Time Warping

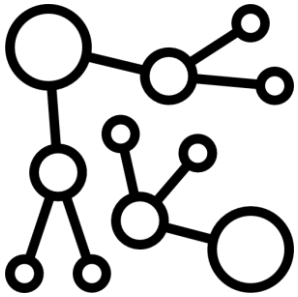


Euclidean Distance



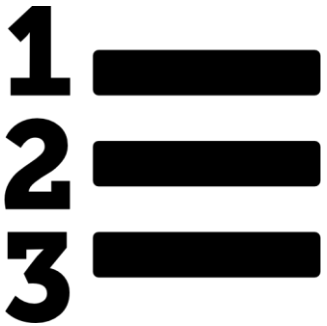
DTW Distance

# Time Series Clustering



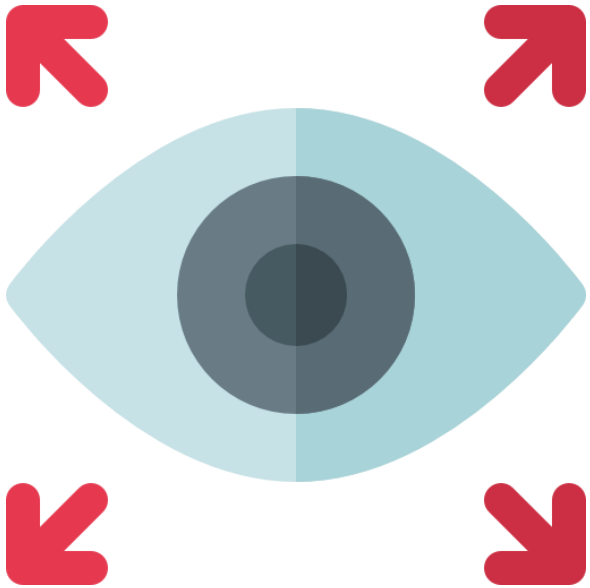
Clustering algorithm

Distance measurement method

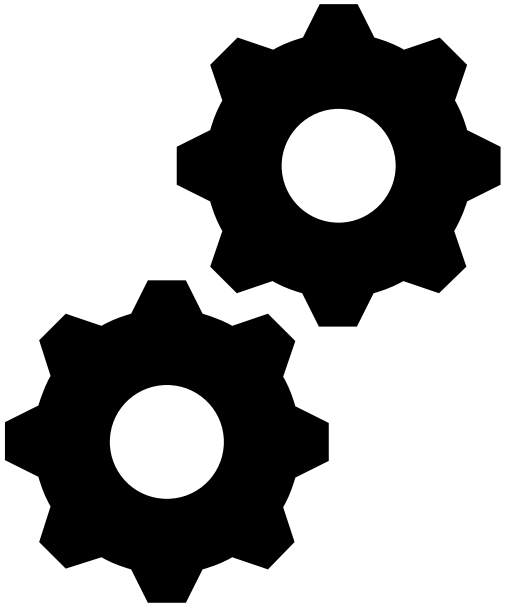



Number of clusters

# Number of Clusters: 13

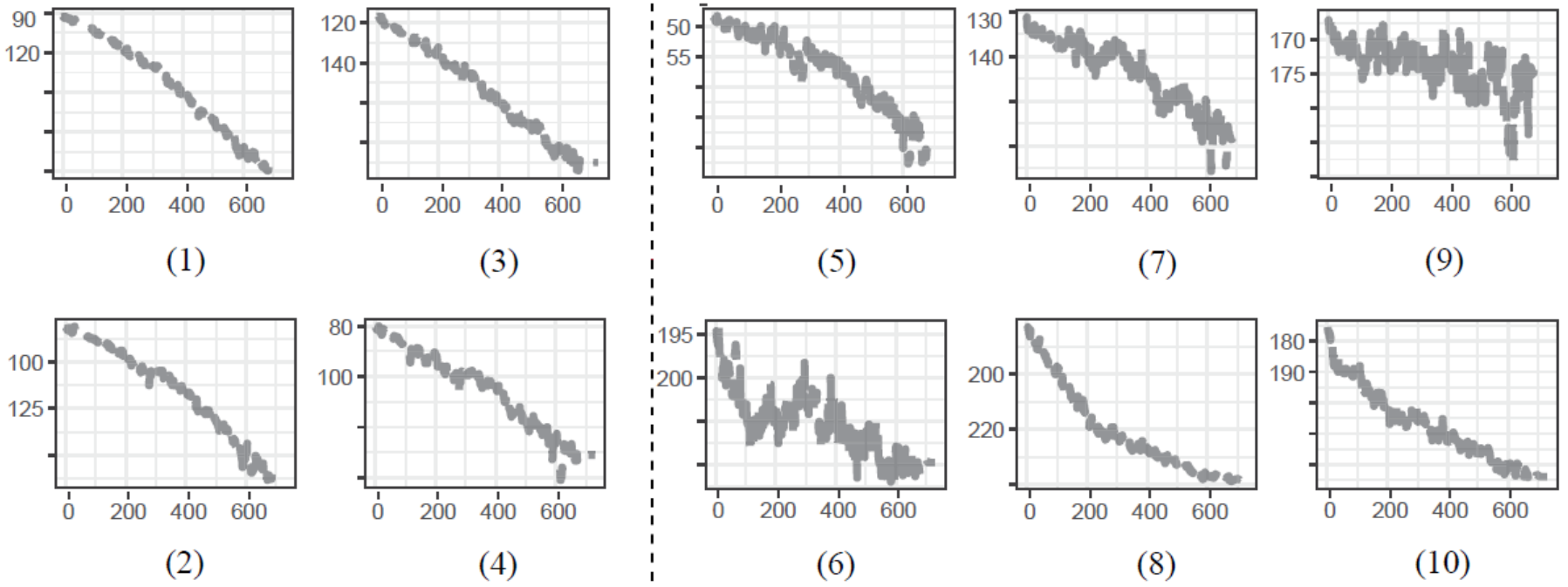


Visual Inspection

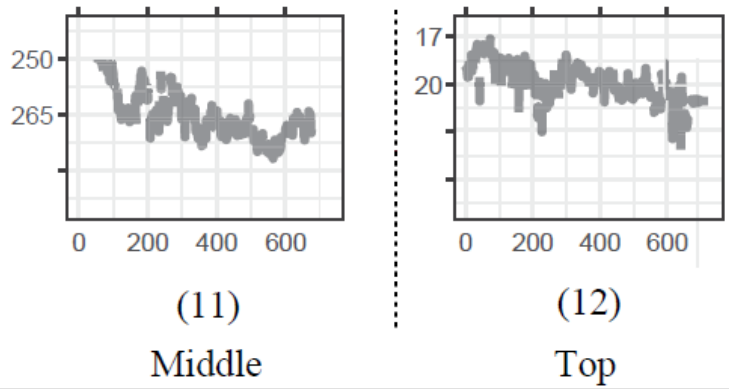


Gap Statistic

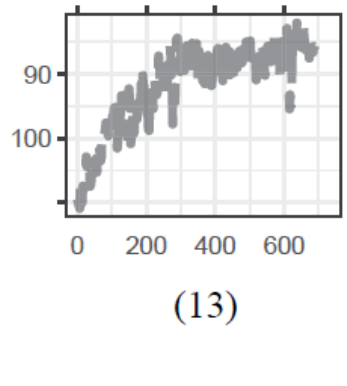
**Falling**

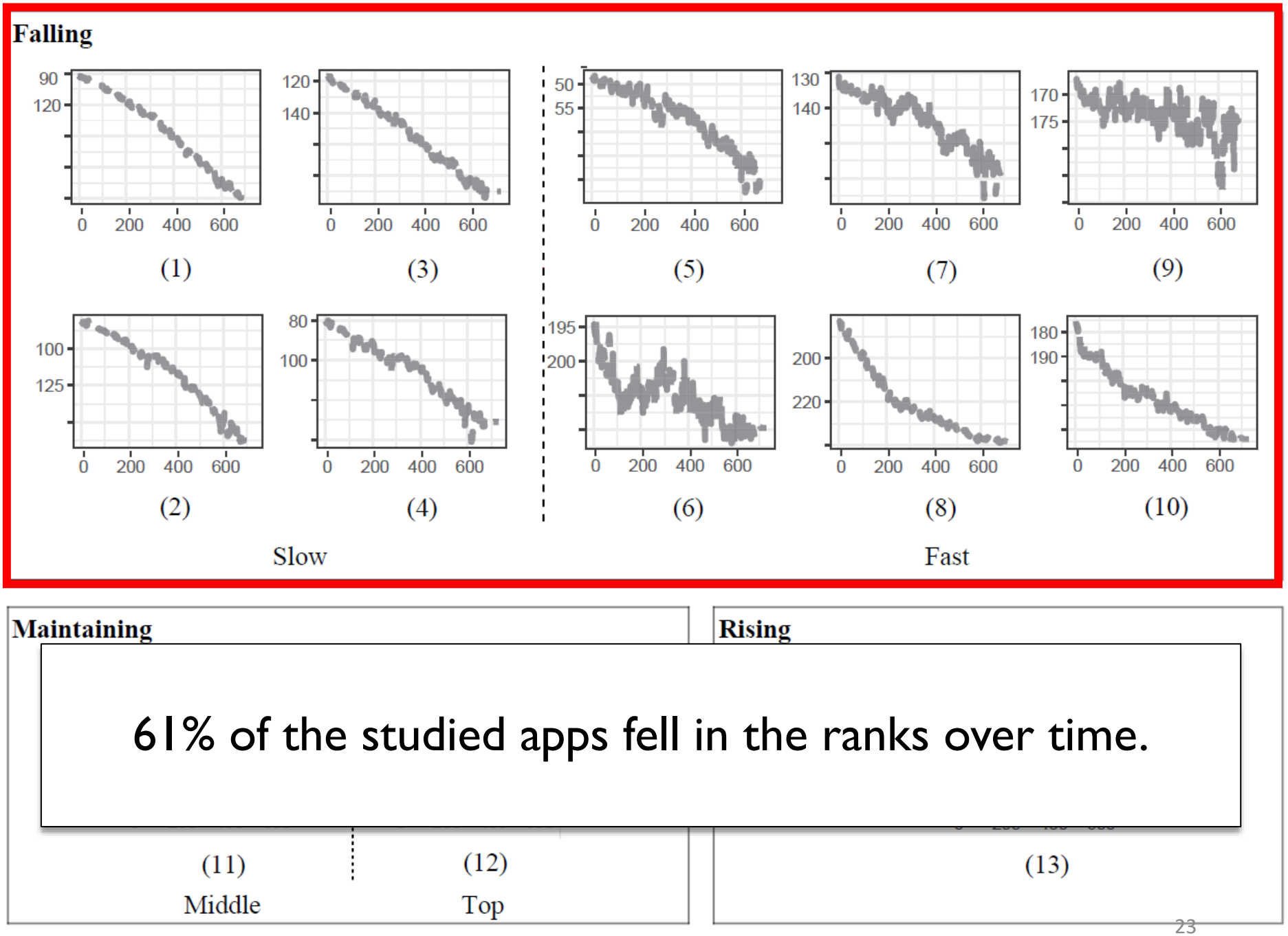



**Maintaining**

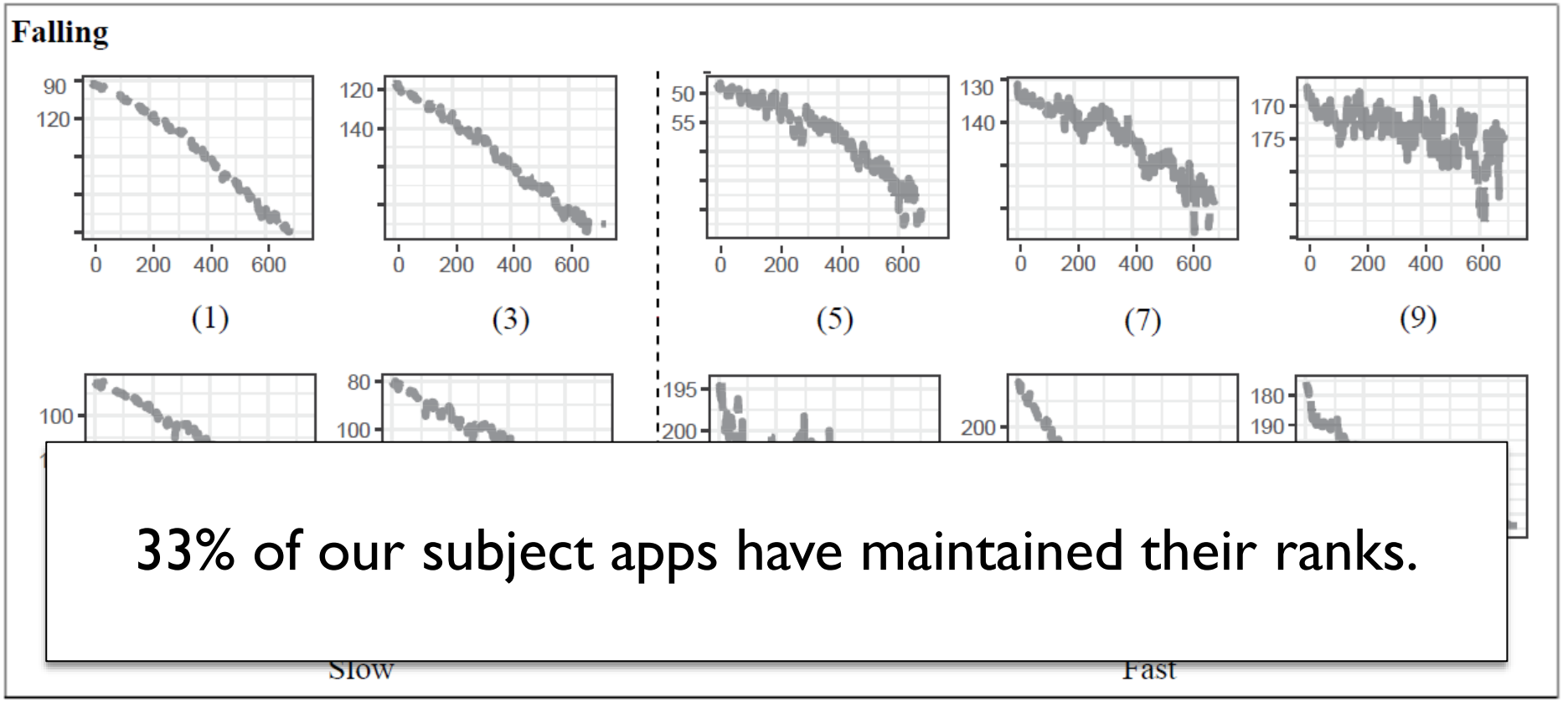



**Rising**

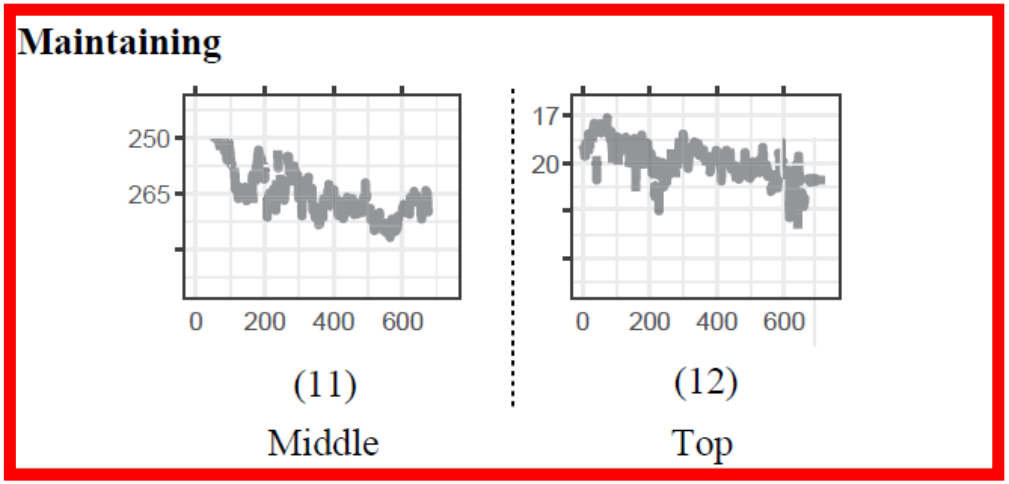


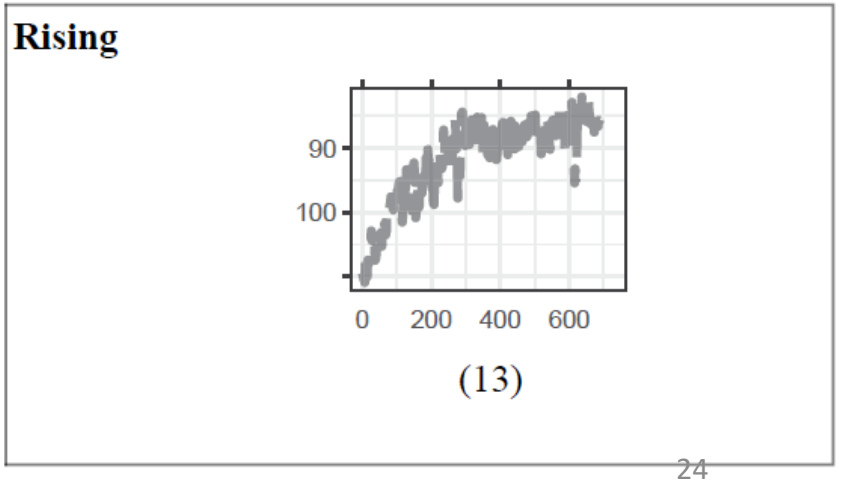


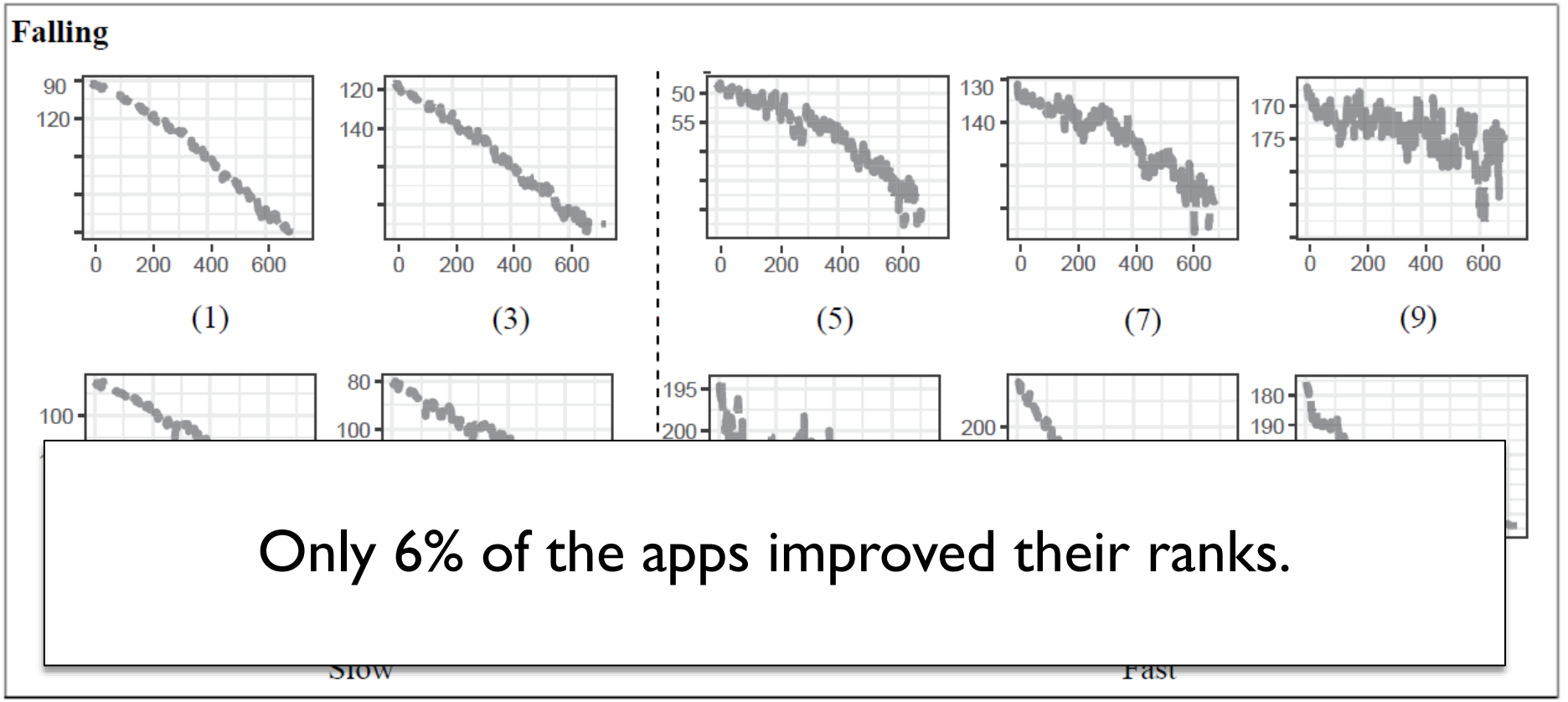

61% of the studied apps fell in the ranks over time.



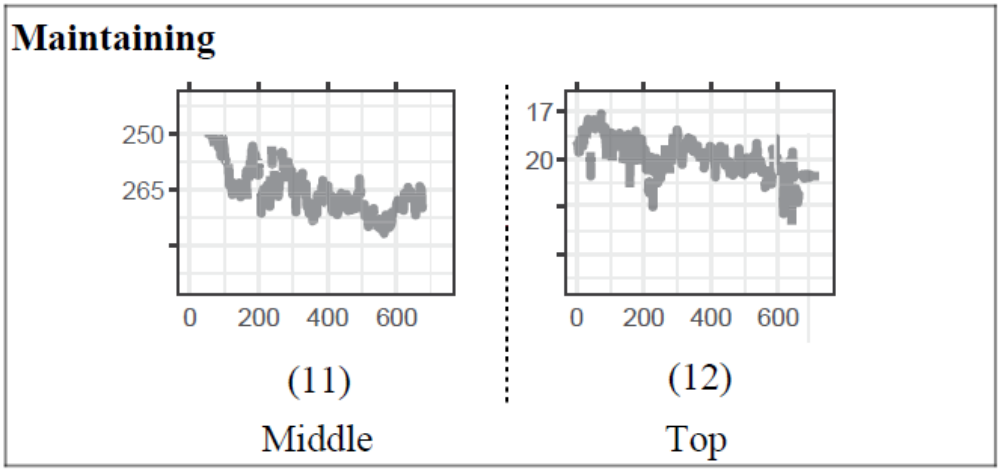

33% of our subject apps have maintained their ranks.

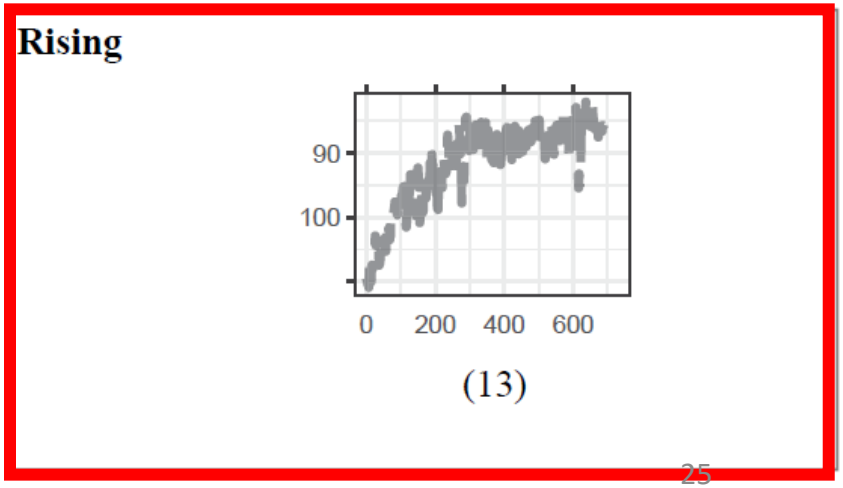


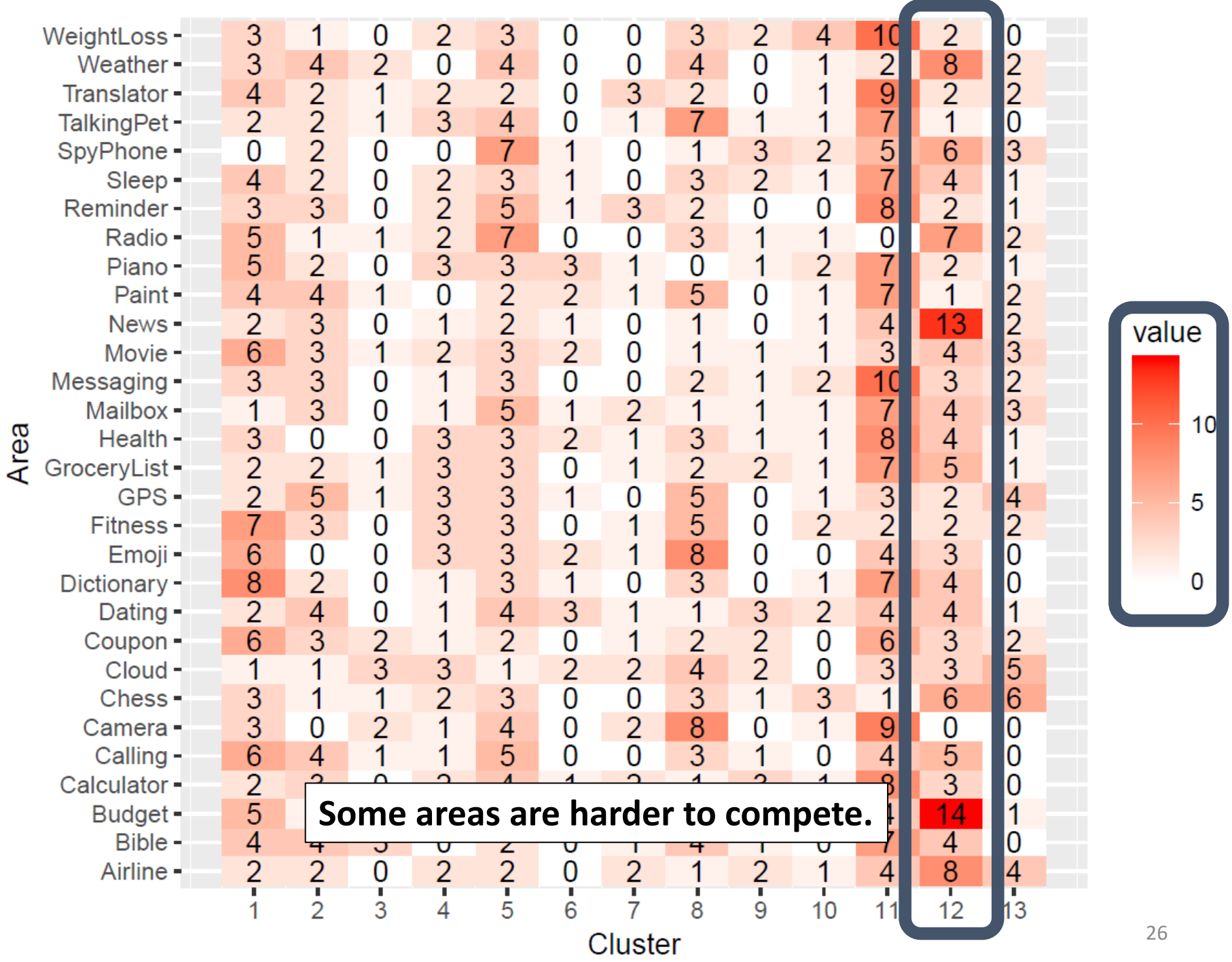





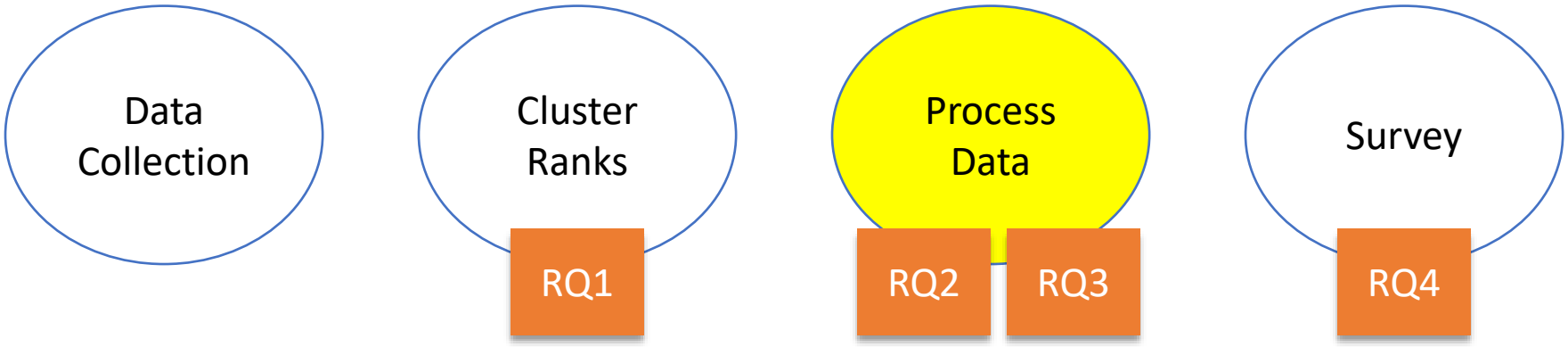

Only 6% of the apps improved their ranks.




Area
WeightLoss 3 1 0 2 3 0 0 3 2 4 10 2 0
Weather 3 4 2 0 4 0 0 4 0 1 2 8 2
Translator 4 2 1 2 2 0 3 2 0 1 9 2 2
TalkingPet 2 2 1 3 4 0 1 7 1 1 7 1 0
SpyPhone 0 2 0 0 7 1 0 1 3 2 5 6 3
Sleep 4 2 0 2 3 1 0 3 2 1 7 4 1
Reminder 3 3 0 2 5 1 3 2 0 0 8 2 1
Radio 5 1 1 2 7 0 0 3 1 1 0 7 2
Piano 5 2 0 3 3 3 1 0 1 2 7 2 1
Paint 4 4 1 0 2 2 1 5 0 1 7 1 2
News 2 3 0 1 2 1 0 1 0 1 4 13 2
Movie 6 3 1 2 3 2 0 1 1 1 3 4 3
Messaging 3 3 0 1 3 0 0 2 1 2 10 3 2
Mailbox 1 3 0 1 5 1 2 1 1 1 7 4 3
Health 3 0 0 3 3 2 1 3 1 1 8 4 1
GroceryList 2 2 1 3 3 0 1 2 2 1 7 5 1
GPS 2 5 1 3 3 1 0 5 0 1 3 2 4
Fitness 7 3 0 3 3 0 1 5 0 2 2 2 2
Emoji 6 0 0 3 3 2 1 8 0 0 4 3 0
Dictionary 8 2 0 1 3 1 0 3 0 1 7 4 0
Dating 2 4 0 1 4 3 1 1 3 2 4 4 1
Coupon 6 3 2 1 2 0 1 2 2 0 6 3 2
Cloud 1 1 3 3 1 2 2 4 2 0 3 3 5
Chess 3 1 1 2 3 0 0 3 1 3 1 6 6
Camera 3 0 2 1 4 0 2 8 0 1 9 0 0
Calling 6 4 1 1 5 0 0 3 1 0 4 5 0
Calculator 2 3 0
Budget 5 14 1
Bible 4 4 0
Airline 2 2 0 2 2 0 2 1 2 1 4 8 4
1 2 3 4 5 6 7 8 9 10 11 12 13
Cluster
Some areas are harder to compete.
value
10
5
0

# Study Setup

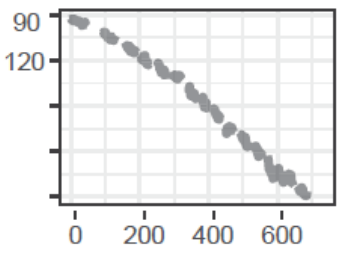

Data Collection
Cluster Ranks
RQ1
Process Data
RQ2
RQ3
Survey
RQ4

# RQ2: Which variables can improve the rank trends?

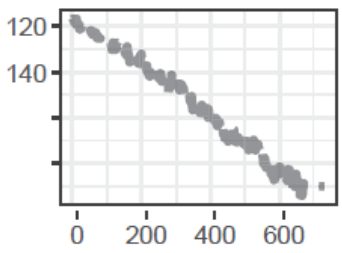
- Optimistic
  - Maintained ranks on Top
  - Rising ranks

- Pessimistic
  - Maintained ranks in the Middle
  - Falling ranks

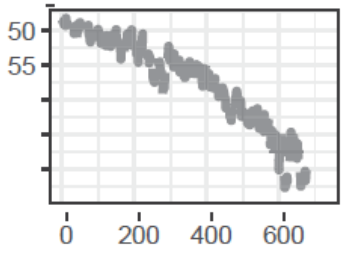
**Falling**



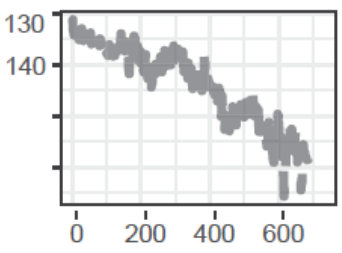

(1)



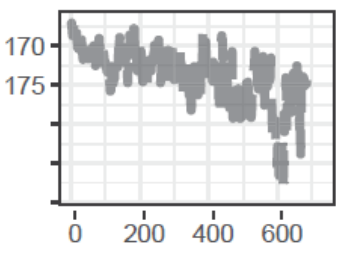

(3)



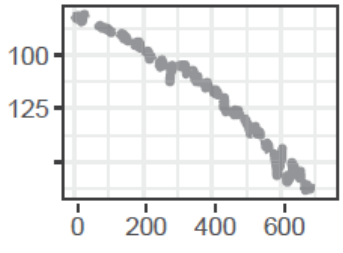

(5)



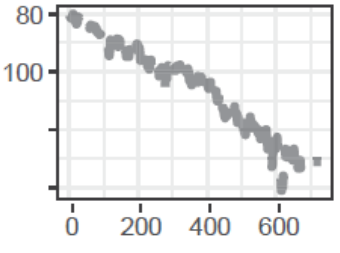

(7)



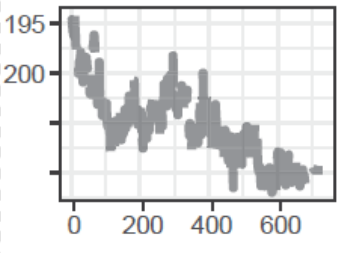

(9)



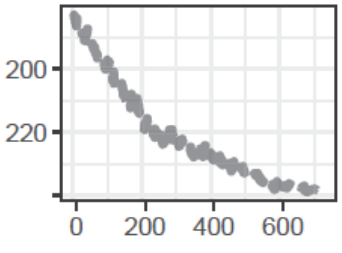

(2)



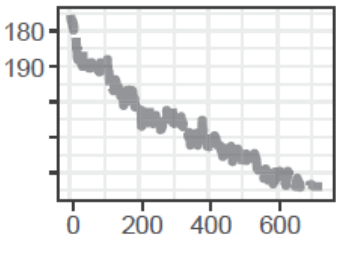

(4)




(6)




(8)

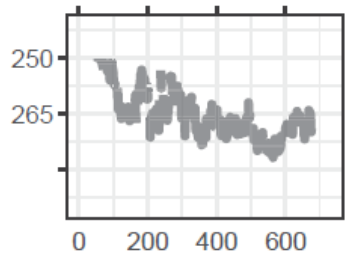



(10)

Slow

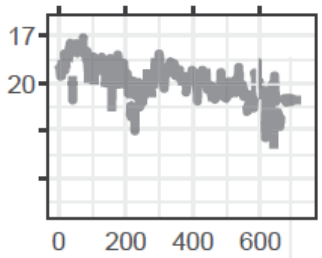
## Optimistic Trends

**Maintaining**




(11)

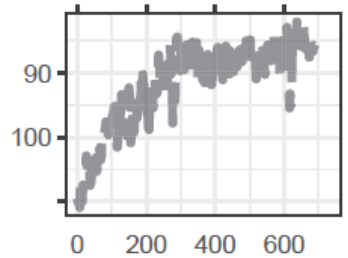
Middle




(12)

Top

**Rising**




(13)

# Independent Variables

#pictures

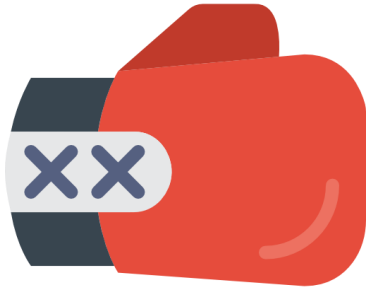
#releases

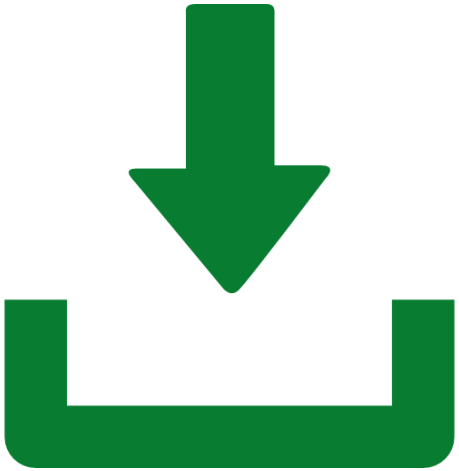
#reviews

Launch D.

# Correlation Analysis

Correlated!





# reviews

# downloads

# Logistic Regression Model

- Pessimistic/Optimistic ~ Independent Variable

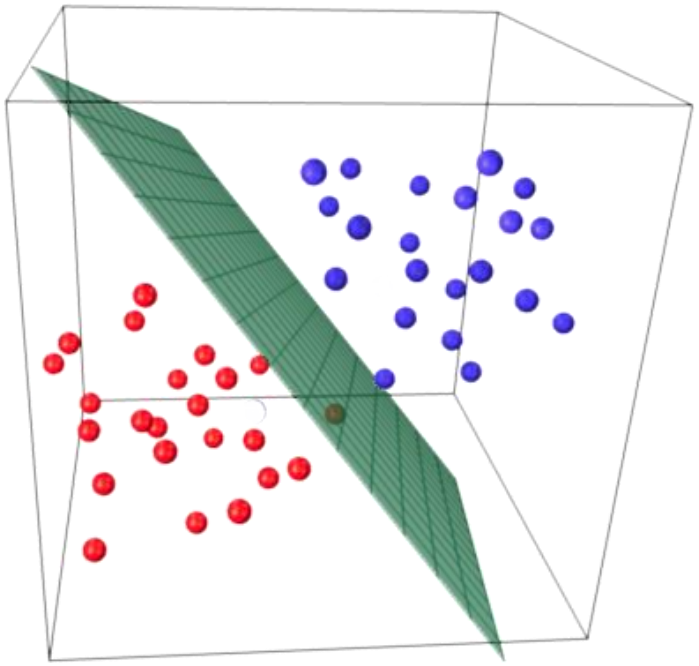
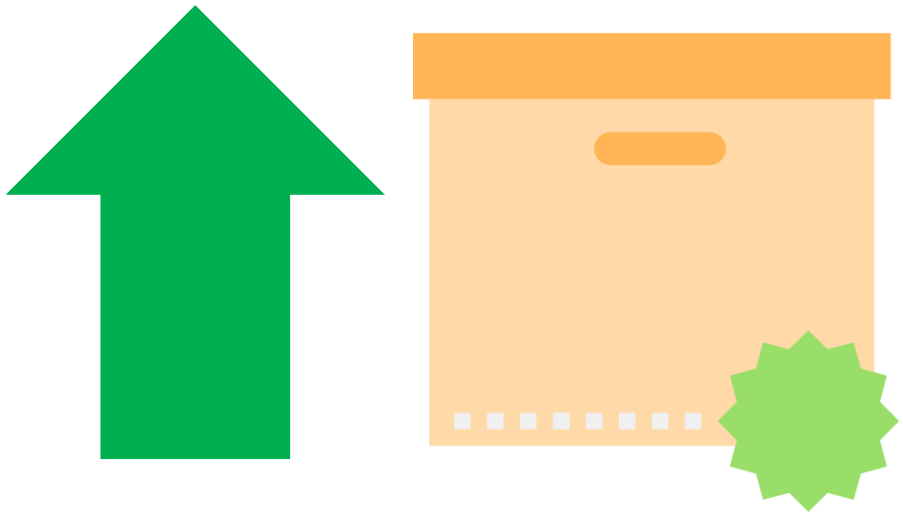


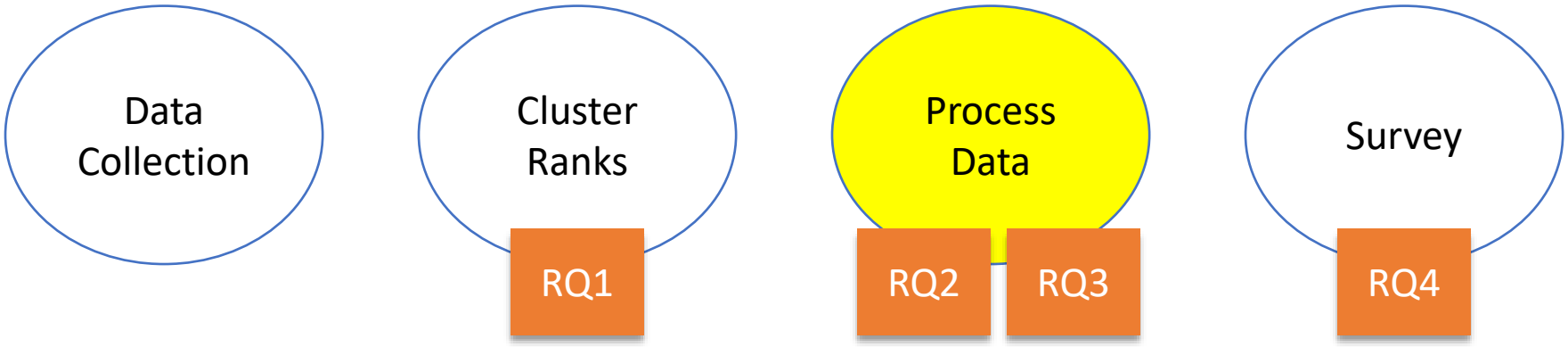
Image Credit: https://codesachin.wordpress.com/2015/08/16/logistic-regression-for-dummies/

Newcomers should be hopeful, since apps that were lunched later are associated with optimistic rank trends.

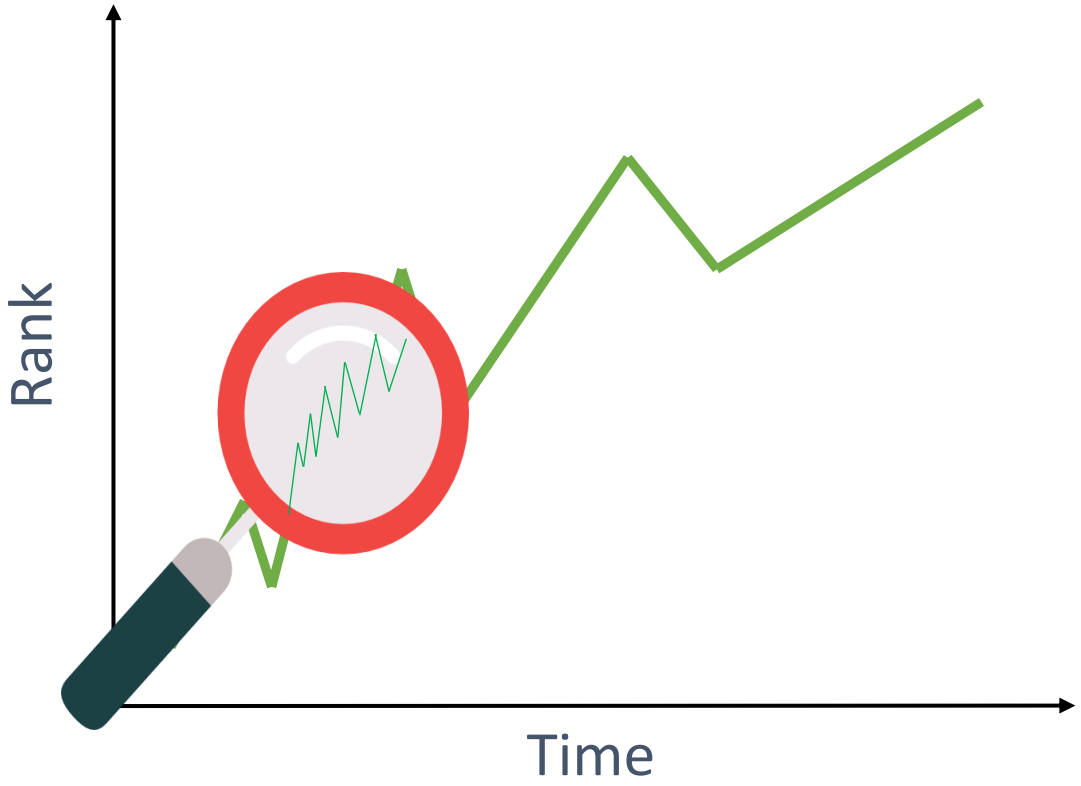
# More releases are encouraged.



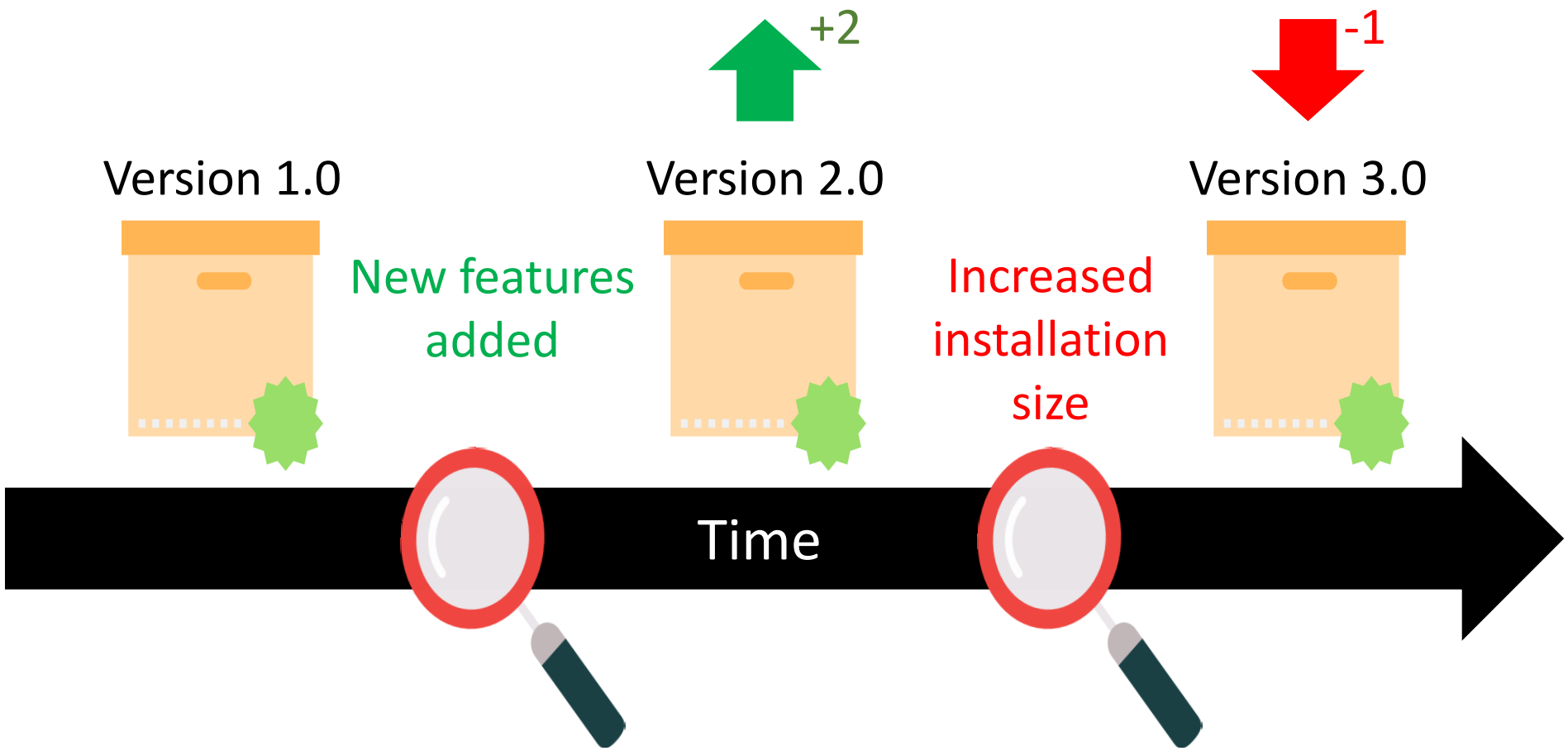
# Study Setup


Data Collection
Cluster Ranks
RQ1
Process Data
RQ2
RQ3
Survey
RQ4

# RQ3: What variables share a significant relationship with the changes in the ranks overtime?



# Example


+2
-1
Version 1.0
Version 2.0
Version 3.0
New features added
Increased installation size
Time

# Mixed Effects Model

Goodness of Fit

Marginal $R^2$

0.02

Conditional $R^2$

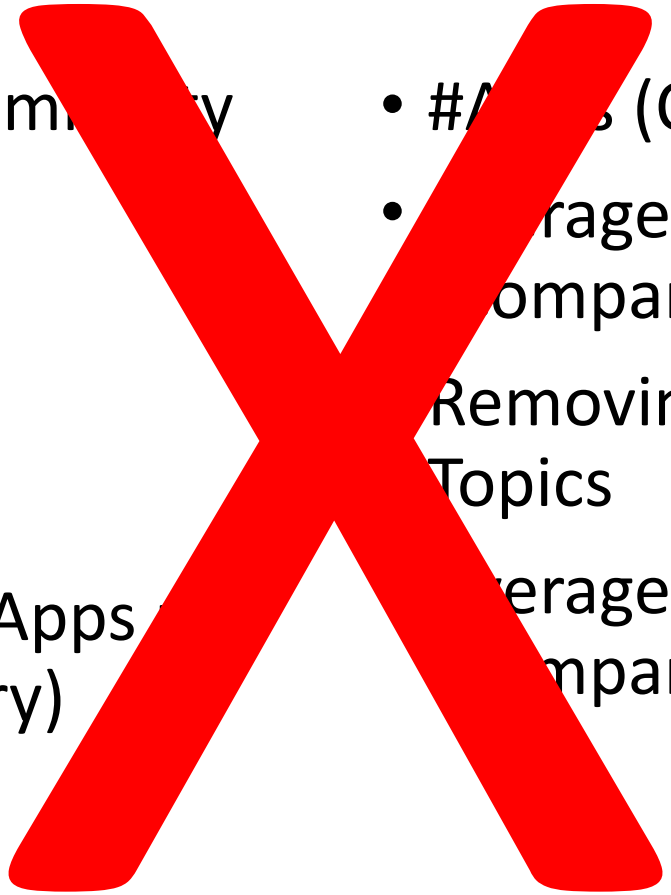
0.68

# Statistically Significant Metrics

- Description Similarity
- Average Price (Category)
- #Pictures
- Name Size
- Ratio of Paid Apps per Total (Category)
- #Apps (Category)
- Average Star-Rating (Company)
- Removing Previous Topics
- Average Price (Company)

# Statistically Significant Metrics

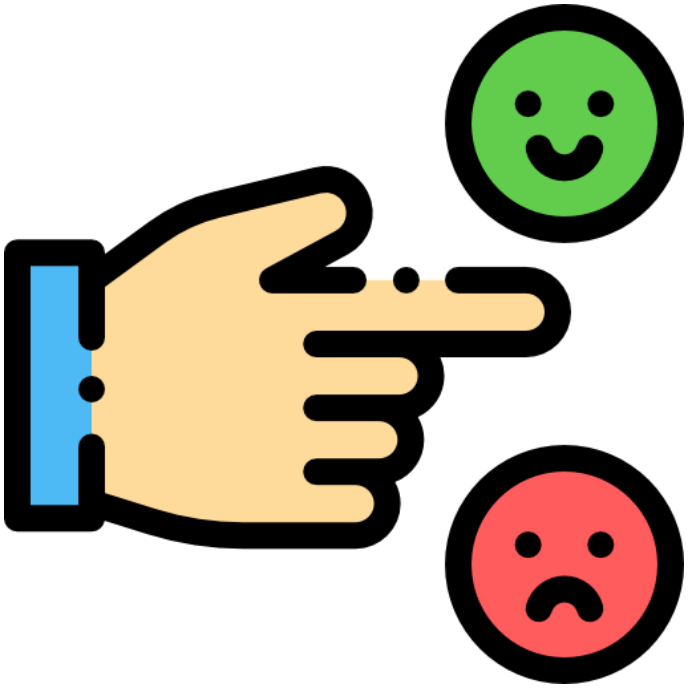
- Description Sim[illegible]ty
- Average Price (Category)
- #Pictures
- Name Size
- Ratio of Paid Apps [illegible] Total (Category)
- #A[illegible]s (Category)
- [illegible]rage Star-Rating [illegible]ompany)
- Removing Previous [illegible]Topics
- [illegible]erage Price [illegible]mpany)



# Common-wisdom does not hold.

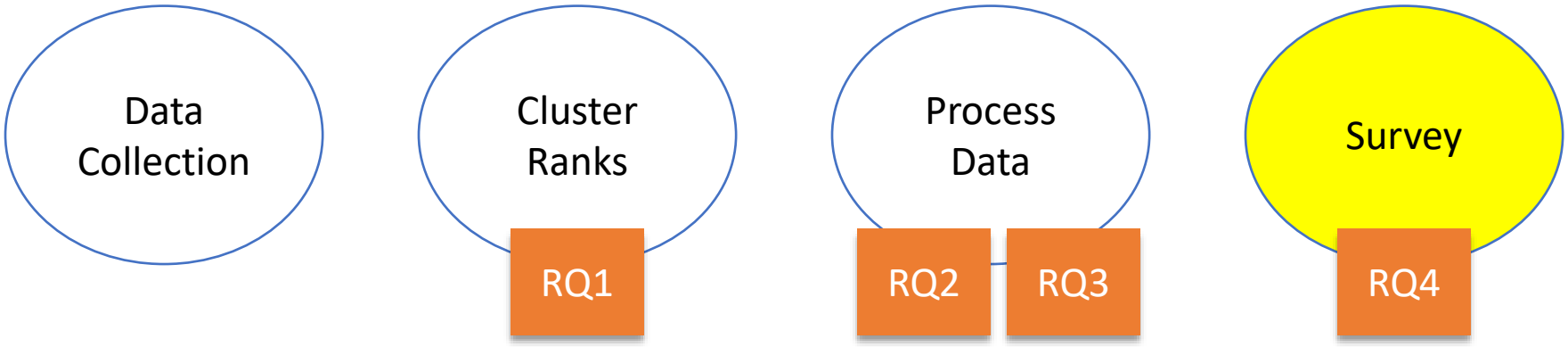
Earth was once believed
to be flat

# User reviews' sentiment score have a strong association with the ranks.



# Developers should not wait too long to publish an update.

# Study Setup


Data Collection
Cluster Ranks
RQ1
Process Data
RQ2
RQ3
Survey
RQ4

*"I wish I could have read this paper earlier!"*

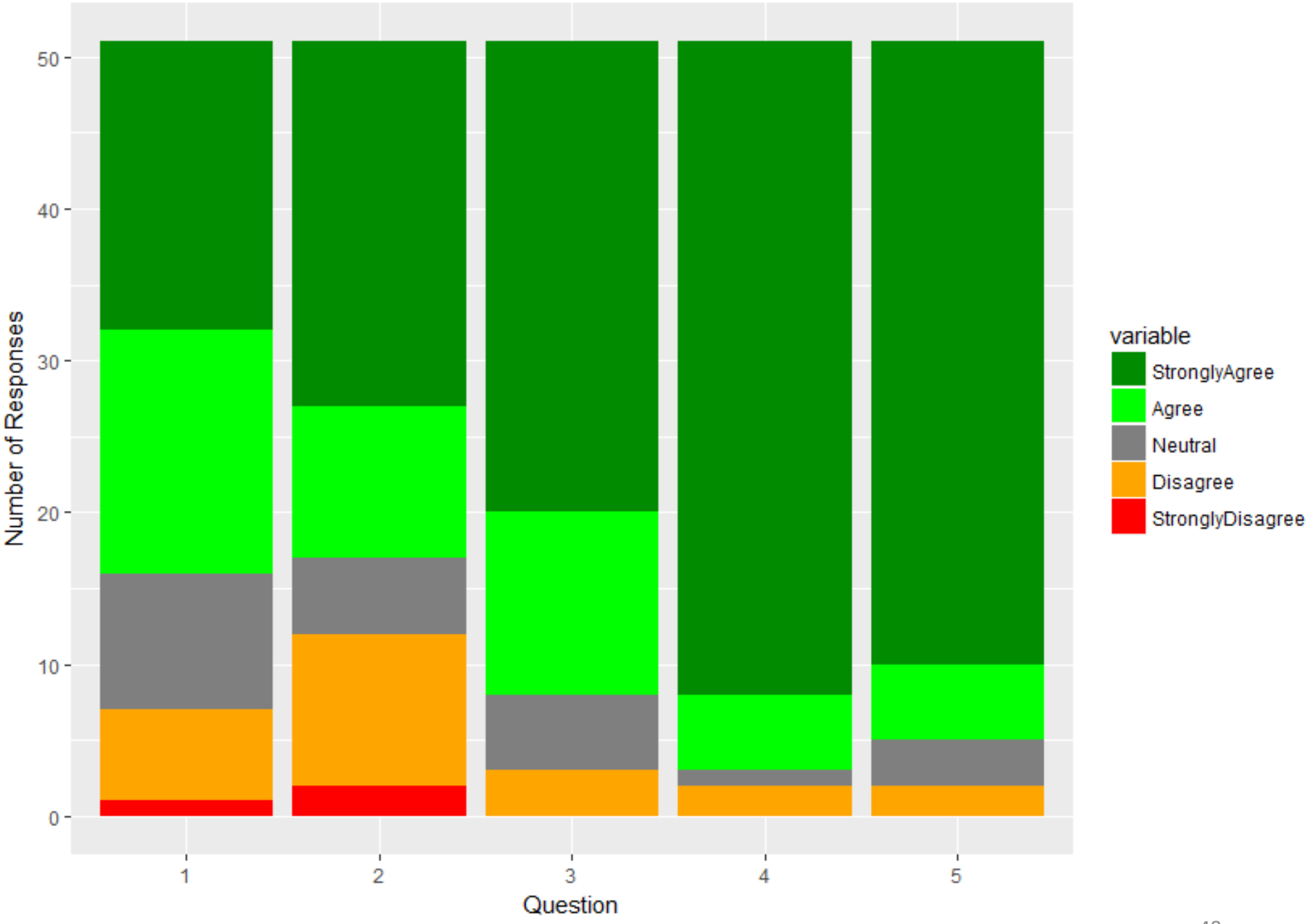
*"It is a very informative paper. It would definitely help app developers."*

*"I confirm your findings and hope developers could better plan for their future according to your work."*

Image Credit: https://www.facebook.com/celebsg4llery/

# Survey of 51 App Developers

1. Do you think that the studied areas are the top areas of mobile apps?
2. Are the considered variables reasonable and sufficient for explaining ranks and rank trends?
3. Do you think the findings of RQ1 are useful and practical for the industry?
4. Do you think the findings of RQ2 are useful and practical for the industry?
5. Do you think the findings of RQ3 are useful and practical for the industry?

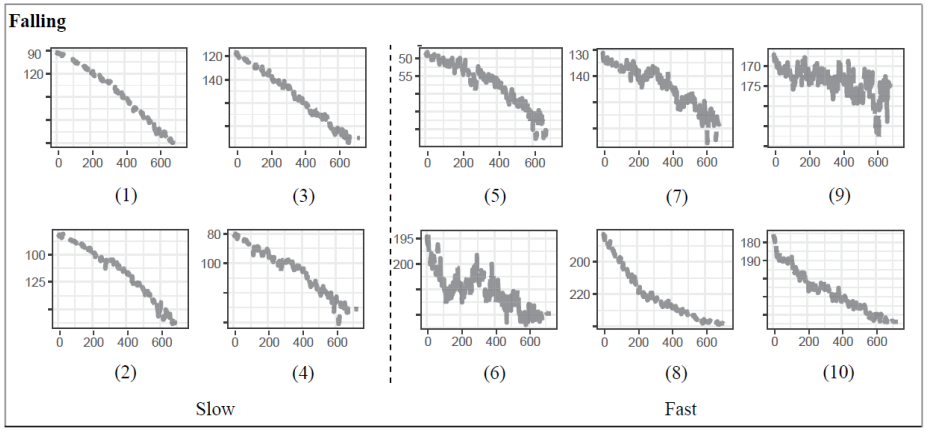

Number of Responses
50
40
30
20
10
0
1
2
3
4
5
Question
variable
StronglyAgree
Agree
Neutral
Disagree
StronglyDisagree

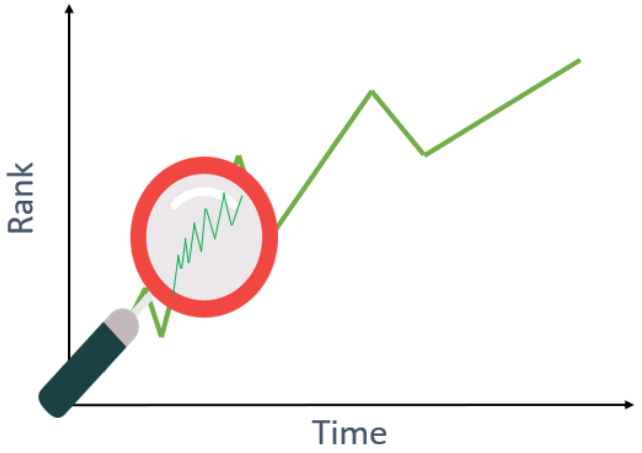
GGR-044

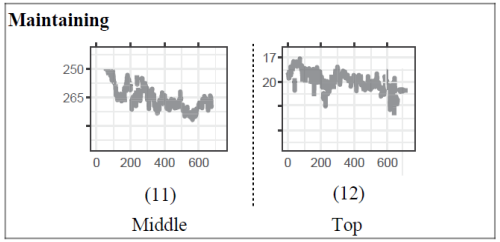
M EV 8842E

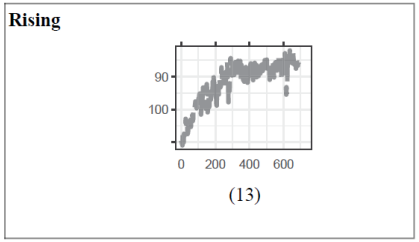
# Implications For New Developers

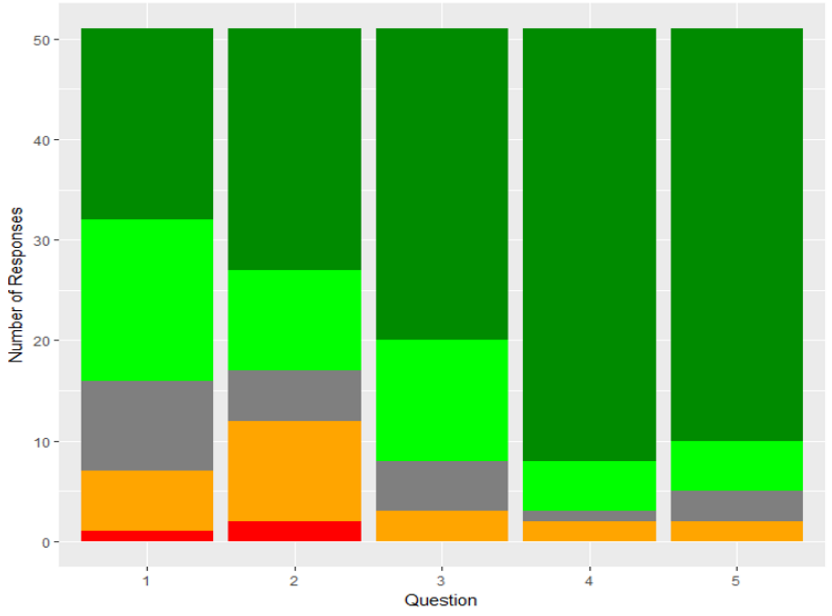
- Developers should carefully select the area in which they intend to work.
- Never is late
  - as in the past two years, many apps have achieved higher ranks in the Google Play Store.
- Investing time and money on the user-reviews
  - Own app
  - Competitors
- Developers should not be distracted by the variables that sound critical but are not much important.




Newcomers should be hopeful, since apps that were lunched later are associated with optimistic rank trends.