

# A Dataset of Open Source Android Applications

Daniel E. Krutz, Mehdi Mirakhorli, and Samuel A. Malachowsky

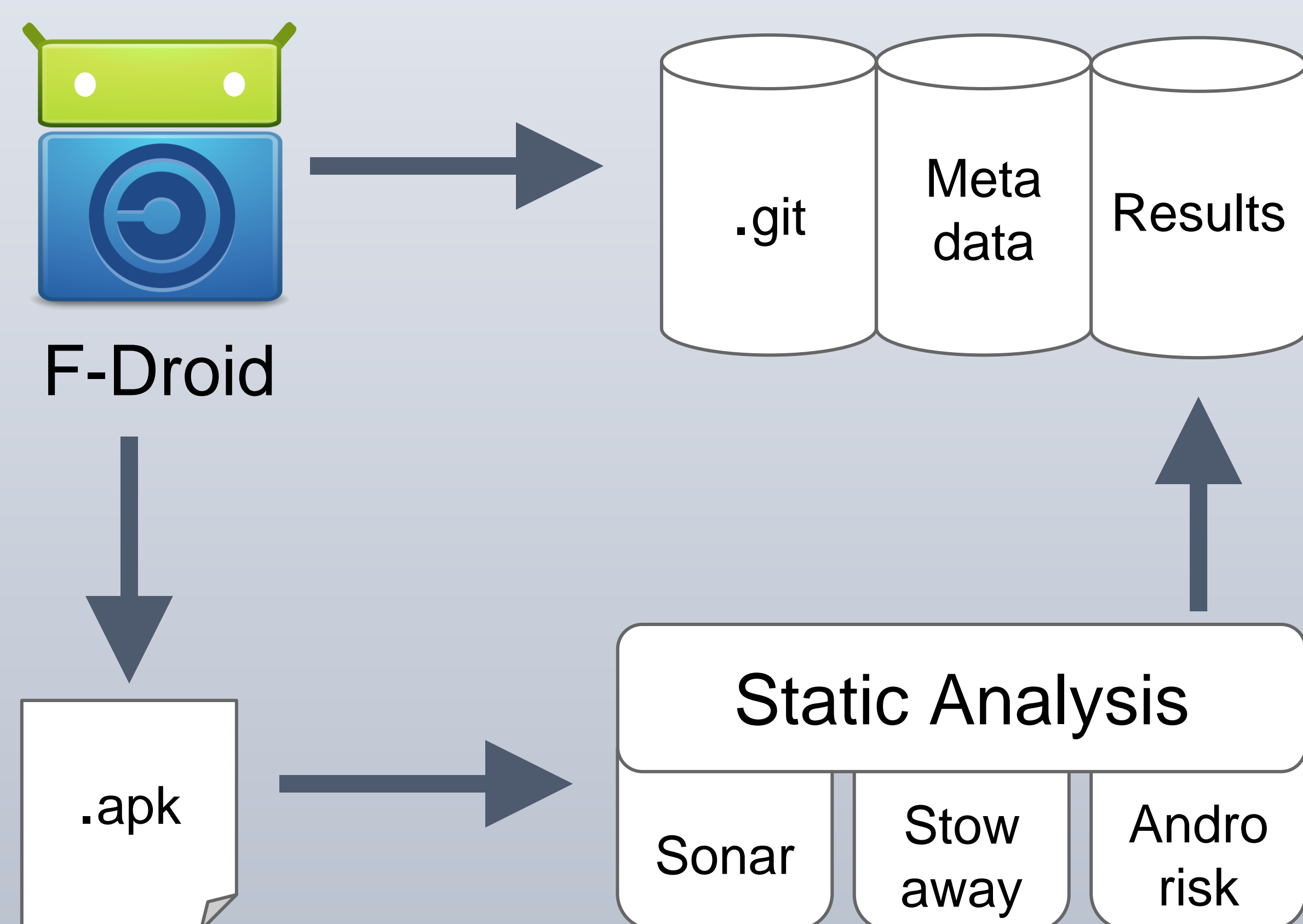
<http://androsec.rit.edu>

## INTRODUCTION

- Began by collecting open source Android apps from F-Droid
- Analyzed Git Histories and performed static analysis using
  - Androrisk: App vulnerability level
  - Sonar: Coding metrics
  - Stowaway: Over & Under Privilege level
- Collected 1,179 apps with 4,416 versions
- 435,680 Version Control commits
- Data is publicly available on website and through SQLite Database

## APPROACH

- Collected data from F-Droid, open source Android website (APK & Git Info)
- Analyzed using various static analysis tools



Semantically Organized Analytical Data

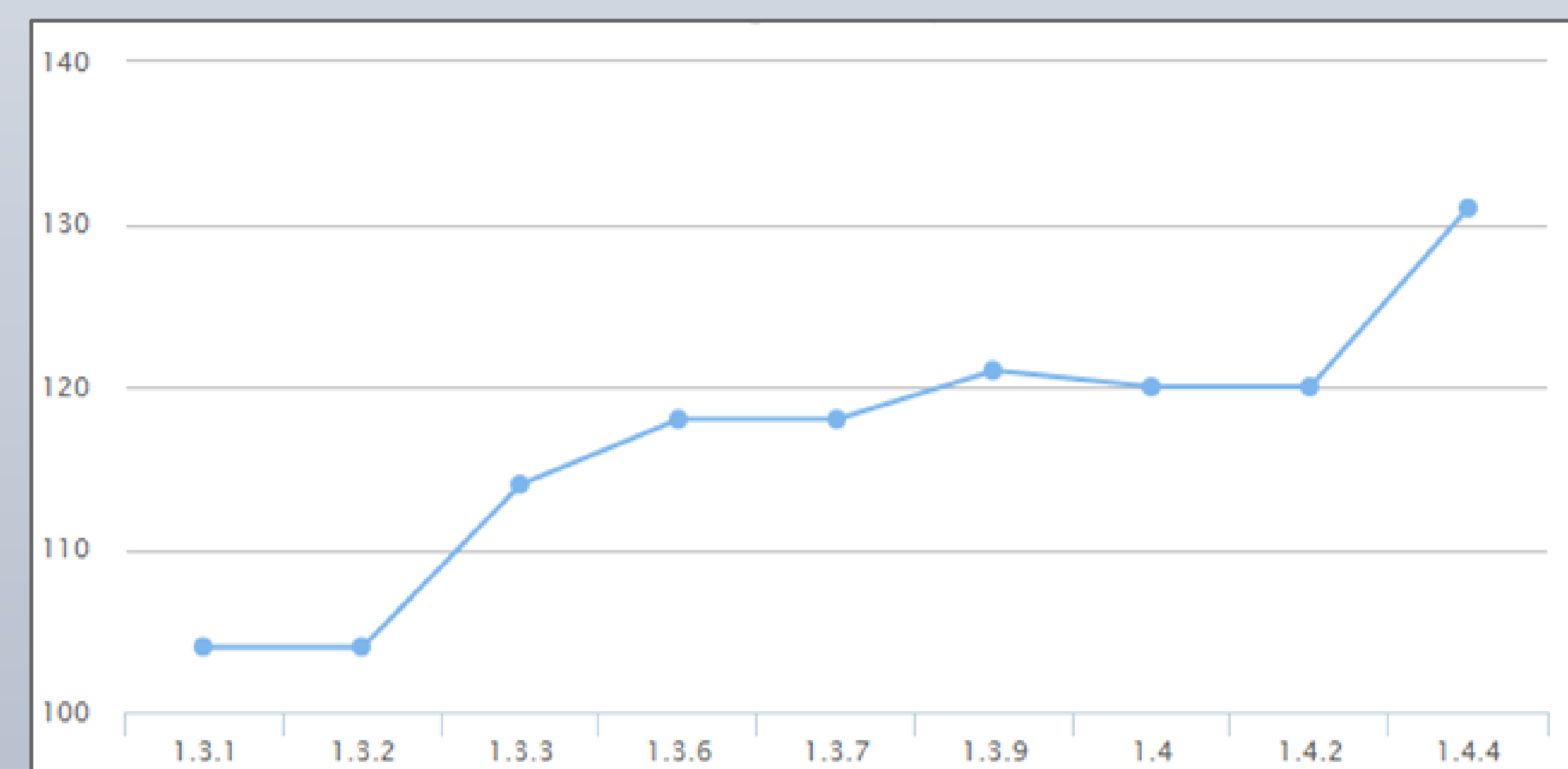
## WEBSITE

Show 10 entries Search: wifi

Name	# Versions	Current Version	Repo Type
WiGLE Wifi Wardriving	9	2.5	git
Better Wifi on/off	1	2.1.0.0	git
Open WiFi Cleaner	4	0.1.6.3	git
WiFi Automatic	9	1.4.4	git
MAXS Module WifiAccess	6	0.0.1.18	srclib
Wifi Fixer	20	1.0.3.5	git

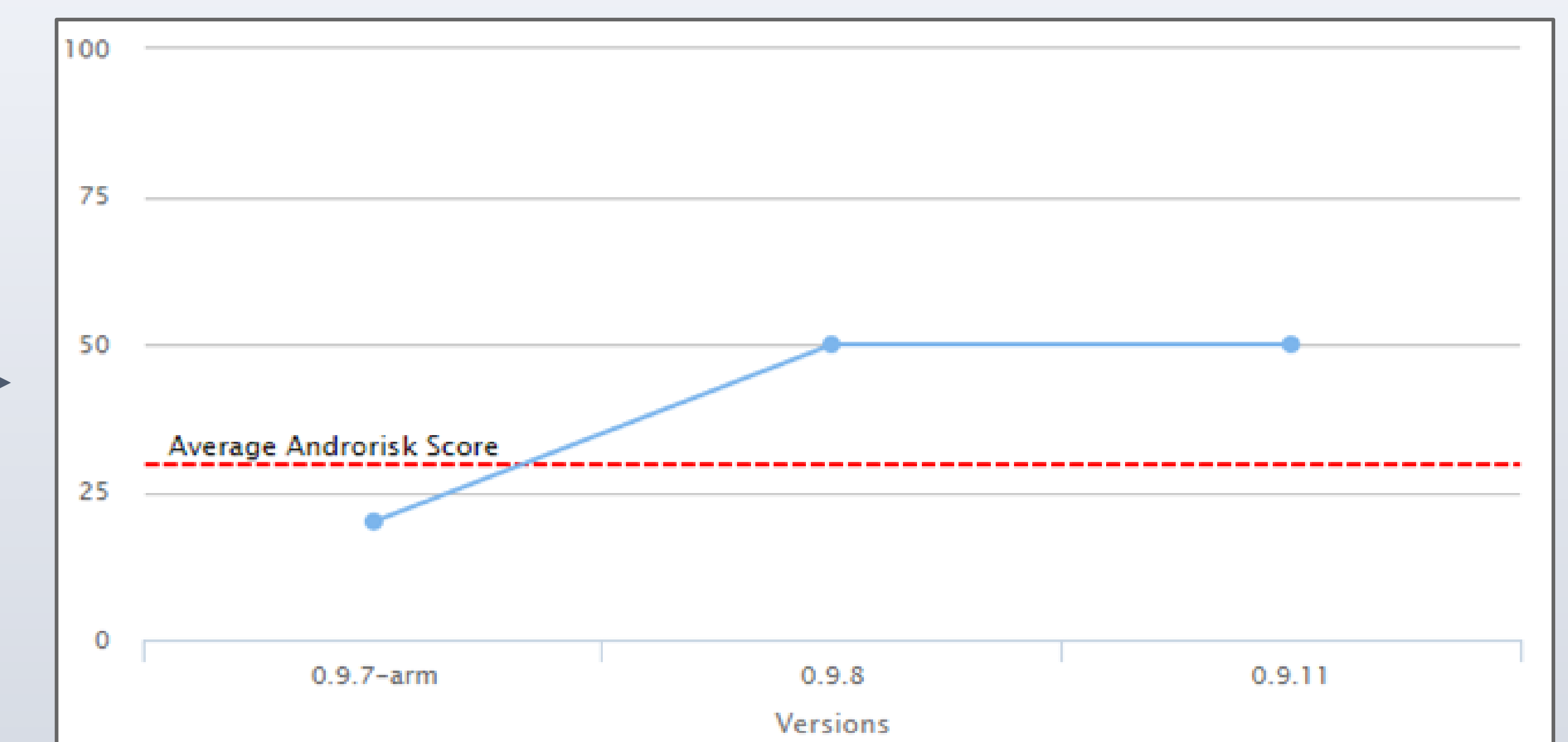
Previous 1 2 Next

Search for app 'FBReader'



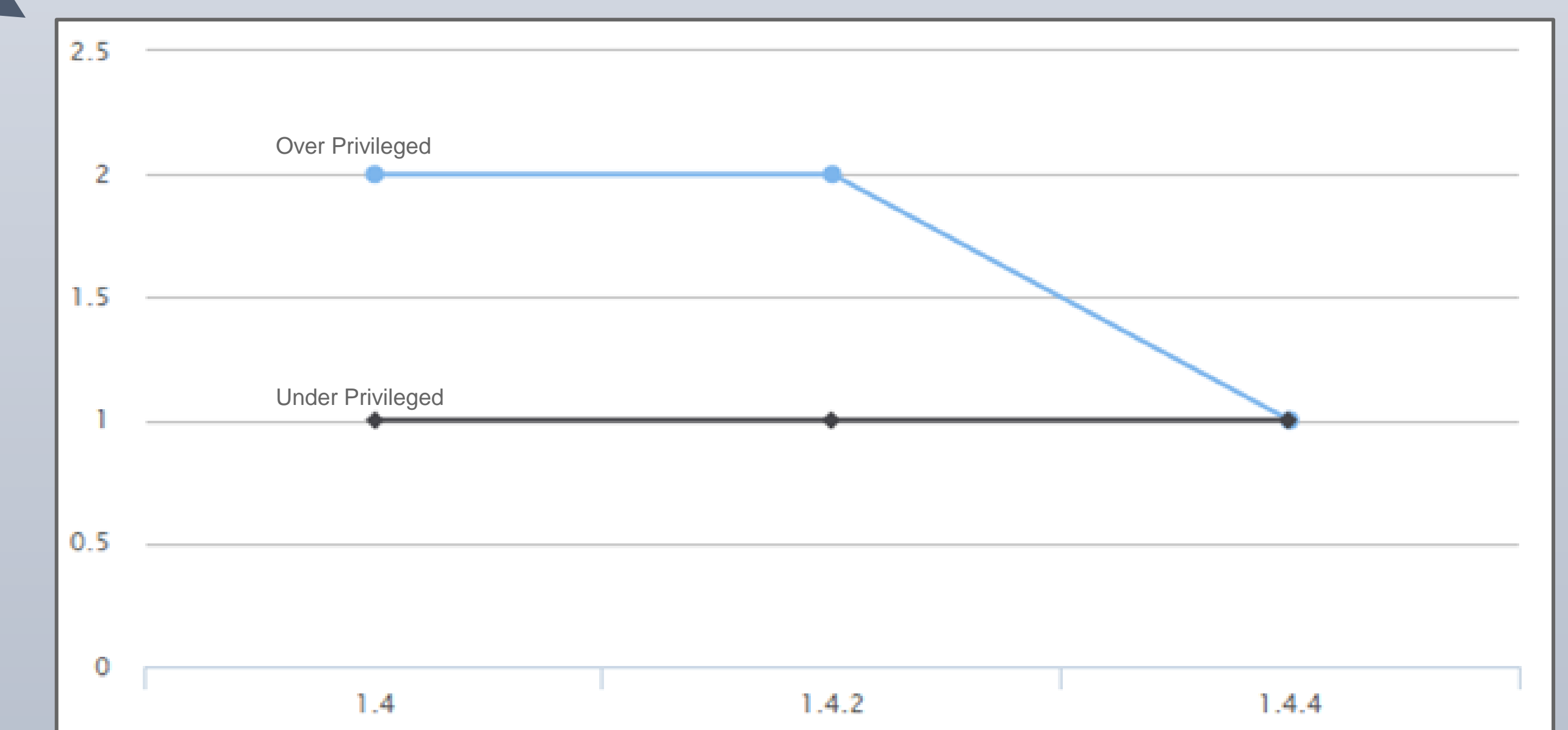
Violations per KLOC - Per Version

Average Violations/KLOC - 113



Androrisk Scores

Average Androrisk Score - 29.66



Over & Under Privileged Rate

Under Privileged Rate- 1.66, Average Over Privileged - 0.66

## DATA

	Value	Count
Totals	Apps	1,179
	Versions	4,416
	Committers	4,535
	Commits	435,680
Largest	Versions	48
	Committers	202
	Commits	65,110

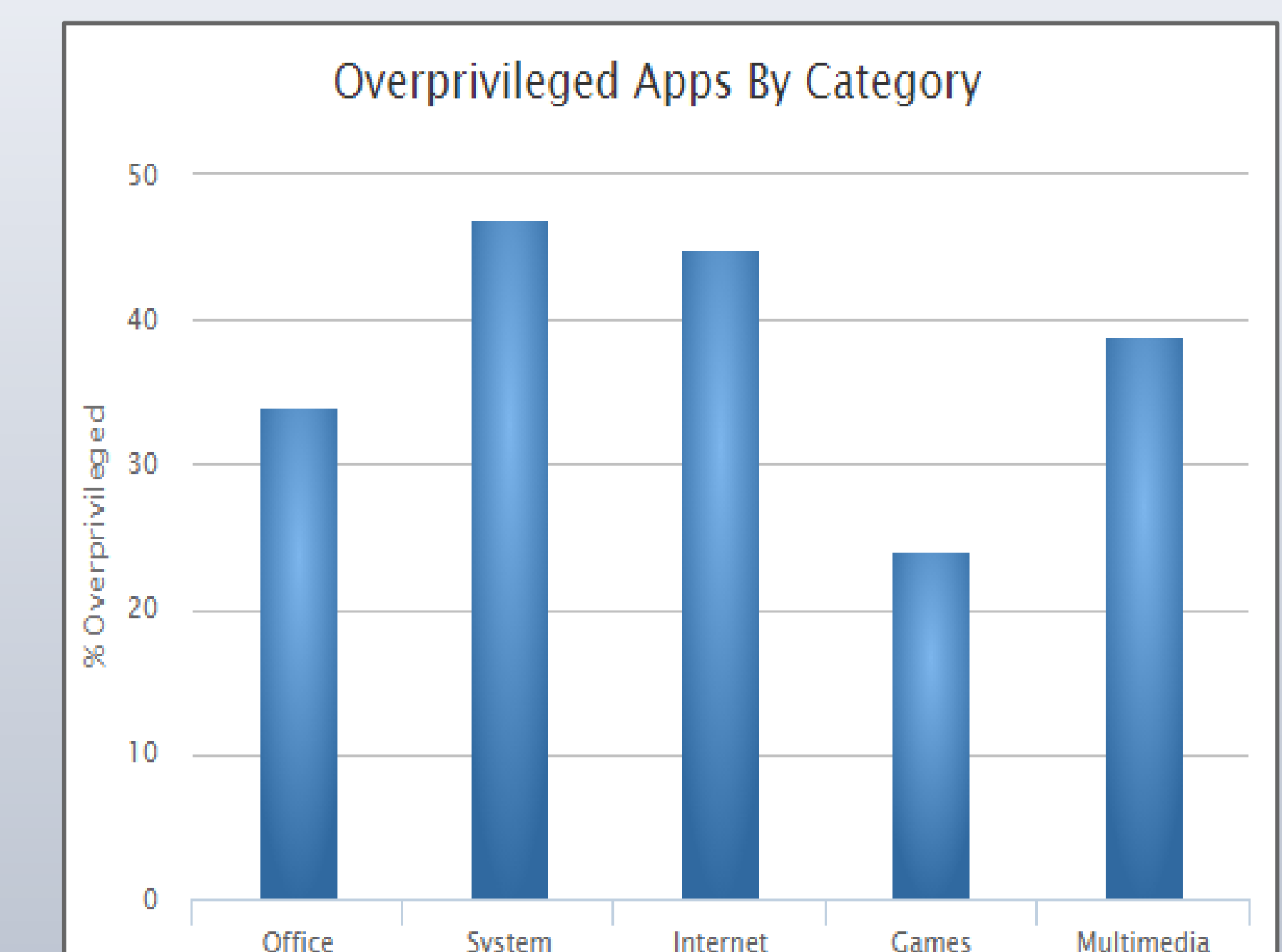
Overall Stats

Query:

select \* from table

Run

Write Your Own Query



Aggregate Statistics