Kurdish Studies
January 2020
Volume: 08, No: 01, pp 254-264
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

DOI: 10.67145/ks.v8i1.4133

A Descriptive and Correlational Analysis of Popularity Patterns Among Trending Google Chrome Extensions

Luis Eduardo Muñoz Guerrero[1*]
[1*]Universidad Tecnológica de Pereira. Email: lemunozg@utp.edu.co.

Abstract
This study presents a descriptive and correlational analysis of a publicly available dataset of trending Google Chrome extensions (N = 50). Browser extensions constitute an increasingly important layer of user-facing software, yet empirical characterizations of what distinguishes popular extensions from the broader marketplace remain scarce relative to the volume of work on extension security and privacy. Using a dataset of extensions flagged as "trending" on the Chrome Web Store, we characterize the distribution of extensions across functional categories, describe the central tendency and dispersion of user ratings and review counts, and examine the association between review volume, category membership, and user ratings through non-parametric statistical tests. Extensions were overwhelmingly concentrated in the Productivity category (58.0% of the sample). Ratings and review counts both departed significantly from normality (Shapiro-Wilk $p < .001$ in both cases), justifying the use of non-parametric methods throughout. Spearman's rank correlation revealed a strong, statistically significant positive association between review count and rating ($\rho = .71$, $p < .001$), indicating that more heavily reviewed extensions in this sample also tended to be more favorably rated. A Kruskal-Wallis test found no statistically significant difference in rating distributions across functional categories ($H(7) = 5.21$, $p = .63$), suggesting that, within this sample, category membership alone was not a strong determinant of user rating. These findings provide a baseline empirical profile of popularity patterns in the Chrome Web Store ecosystem and offer a methodological and substantive reference point for future studies on browser extension adoption, marketplace design, and user trust signals.
Keywords: Chrome extensions, browser extension marketplace, popularity analysis, correlational study, descriptive statistics, non-parametric methods

1. Introduction

Browser extensions have become a core mechanism through which users customize and extend the functionality of modern web browsers. What began as a relatively niche feature aimed at developers and power users has evolved into a mainstream layer of the browsing experience: ad blockers, password managers, productivity tools, shopping assistants, and accessibility utilities are now installed by hundreds of millions of users worldwide. The Google Chrome Web Store, as the primary distribution channel for Chromium-based browsers, hosts a large and heterogeneous ecosystem of extensions spanning productivity, security, entertainment, developer tooling, social media integration, and shopping, among other categories.

Despite the scale and everyday relevance of this ecosystem, relatively little empirical work has characterized the factors associated with extension popularity using publicly available metadata. This is a notable gap when contrasted with the substantial body of research dedicated to browser extension security: numerous studies have examined how extensions request excessive permissions, how malicious extensions evade store review processes, and how vulnerabilities in the extension APIs can be exploited by third parties. By comparison, the question of what makes an extension popular in the first place — independent of whether it poses a security risk — has received comparatively little systematic, data-driven attention. This asymmetry is understandable: security incidents are high-salience events that motivate urgent research, whereas popularity is a more mundane, slower-moving phenomenon. Nonetheless, understanding popularity patterns has practical value. Store operators, extension developers, and researchers studying platform governance all benefit from a clearer empirical picture of which categories tend to accumulate more user engagement, and whether that engagement (measured through review volume) is associated with more favorable or more polarized user sentiment (measured through average rating).

This paper addresses that gap through a descriptive and correlational analysis of a publicly available dataset of trending Google Chrome extensions. Rather than proposing a predictive model or a causal account of what drives popularity, this study aims to establish a baseline empirical characterization of the relationship between extension category, user ratings, and review volume. We deliberately adopt a descriptive and correlational design, rather than a predictive or causal one, for two reasons. First, the dataset used here is a cross-sectional snapshot rather than a longitudinal record, so any association observed at a single point in time could reflect either direction of influence, or neither — a concern we return to directly in the Limitations. Second, we consider it methodologically useful to first establish a clear, well-documented descriptive baseline before more complex modeling efforts are attempted on this or similar datasets — a step that is often skipped in the rush toward predictive modeling, but which provides essential context for interpreting more sophisticated results. The remainder of this paper is organized as follows. Section 2 reviews related work on marketplace popularity studies and browser extension research more broadly, situating the present study within that literature. Section 3 describes the dataset, preprocessing steps, and the statistical methodology used to answer each research question. Section 4 presents the results, organized around the three research questions introduced below. Section 5 discusses the findings in light of the related

work, interprets their practical implications, and outlines the limitations of the study. Section 6 concludes and outlines directions for future work.

To operationalize this gap, the study is guided by three research questions, each corresponding to a distinct but related facet of extension popularity:

RQ1: How are trending Chrome extensions distributed across functional categories? This question is purely descriptive and establishes the compositional baseline against which the other two research questions are interpreted. Understanding the categorical composition of the trending set is a necessary first step, since any observed association between rating and review count could, in principle, be confounded by an uneven distribution of extensions across categories.

RQ2: Is there a statistically significant association between review count and user rating? This question probes whether extensions that attract more user engagement (as proxied by review volume) also tend to be rated more favorably (or unfavorably) by users, or whether the two are effectively independent. A positive association would be consistent with a "popularity breeds satisfaction" or self-reinforcing adoption narrative; a null or negative association would suggest that visibility and satisfaction are largely decoupled in this marketplace.

RQ3: Do user ratings differ significantly across extension categories? This question examines whether the functional category of an extension is itself informative about the ratings it tends to receive, independent of review volume. Significant differences would suggest that user expectations, competitive dynamics, or the nature of the underlying task (e.g., ad blocking versus entertainment) shape rating behavior in category-specific ways.

Taken together, answering these three questions yields the following contributions:

- A descriptive statistical profile of the trending Chrome extensions dataset, including the distribution of extensions across functional categories and the central tendency and dispersion of ratings and review counts.
- A correlational analysis between review volume and user rating using an appropriate non-parametric method (Spearman's ρ), together with an explicit justification of that methodological choice based on the observed distributional properties of the data.
- A category-level comparison of rating distributions using the Kruskal-Wallis test.
- A publicly documented, reproducible analysis pipeline (Python, pandas, scipy) that can serve as a methodological reference for future descriptive studies on marketplace metadata, including software marketplaces beyond the Chrome Web Store.

2. Related Work

Two adjacent bodies of literature inform the present study: work on popularity and success in digital marketplaces more broadly, and work specifically on browser extensions. A substantial body of work has examined what drives popularity, adoption, and perceived quality in digital software marketplaces, most notably the Google Play Store and the Apple App Store. Tian et al. (2015) analyzed 28 factors across eight dimensions to compare high-rated and low-rated free Android applications on Google Play, finding statistically significant differences on 17 of the 28 factors, including app category/genre. Potharaju et al. (2018) conducted a longitudinal analysis of more than 160,000 Google Play apps and found that a small number of developers control a disproportionate share of total downloads, consistent with a highly right-skewed, "hit-driven" distribution of engagement metrics.

Most directly relevant to RQ2, Finkelstein et al. (2017) mined the BlackBerry World App Store and reported strong correlations between customer rating and popularity (operationalized as download rank), together with a statistically significant rating advantage for free apps over paid ones. These studies typically draw on metadata such as app category, download counts, price, and textual review content to characterize or predict app success.

Most relevant to RQ1, Zhong and Michahelles (2013) analyzed a large dataset of Google Play transactions and concluded that the marketplace behaves more like a "superstar" market dominated by a small number of hit products than a "long-tail" market, with blockbuster apps showing both higher download counts and higher ratings than the rest of the marketplace.

This body of work, however, has focused overwhelmingly on mobile application marketplaces rather than browser extensions, which differ in important respects: extensions typically have a narrower, more utilitarian scope than mobile apps, are usually free, and are installed within an existing browsing session rather than sought out as standalone destinations.

Turning from mobile-application marketplaces to browser extensions specifically, the great majority of empirical research here has focused on security and privacy concerns rather than on popularity or adoption per se.

This line of work dates back at least to Carlini, Felt, and Wagner (2012), who conducted a security review of 100 Chrome extensions and identified 70 vulnerabilities across 40 of them, establishing one of the foundational empirical accounts of extension-level security risk.

Notably, the gap between Carlini et al.'s technical risk findings and user-perception findings suggests that user ratings — the dependent variable of interest here — may reflect functional satisfaction more than an informed assessment of security or privacy risk. This is corroborated directly by Aggarwal et al. (2018), who studied "spying" extensions on the Chrome Web Store (i.e., extensions covertly exfiltrating user browsing data) and found that these extensions exhibited a distribution of user counts, ratings, and review counts statistically similar to that of the broader population of benign extensions — meaning popularity metrics alone offer users essentially no signal for distinguishing spying extensions from ordinary ones. This finding supports treating rating and review count, as done here, strictly as measures of perceived popularity and satisfaction rather than as proxies for security or quality. The present study is positioned explicitly in this less-studied space: rather than asking which extensions are dangerous, it asks which extensions are popular, and whether that popularity correlates with more favorable user sentiment.

Beyond substantive findings, prior work also offers methodological precedent for the analytical approach adopted here. Kitchenham et al. (2016) provide a large-scale, empirical-software-engineering-specific treatment of when non-parametric methods are appropriate, explicitly recommending rank-based tests such as the Kruskal-Wallis test over parametric alternatives (e.g., ANOVA) when working with non-normally distributed software engineering data — precisely the justification underlying the methodological choices adopted here. Such precedents support the preference for rank-based, distribution-free statistical tests over their parametric counterparts when working with skewed count data such as review volume, a preference later confirmed empirically by the non-normality of both variables in this dataset (Table 2).

Building on this literature, the present study applies these methodologically well-precedented descriptive and correlational techniques to a browser extension marketplace dataset, aiming to provide a foundational empirical reference that subsequent, more elaborate studies (e.g., predictive modeling of ratings, longitudinal tracking of extension popularity, or comparative analysis across browser vendors) can build upon.

3. Materials and Methods

We use the "Trending Google Chrome Extensions" dataset, publicly available on Kaggle (Patra, 2019), published on September 20, 2019, under a CC0: Public Domain license. The dataset was selected for three reasons relevant to the goals of this study. First, it is drawn directly from the Chrome Web Store and therefore reflects real marketplace metadata rather than a synthetic or heavily preprocessed derivative. Second, at the time of writing it has attracted minimal prior analytical attention — a single publicly associated exploratory notebook and no indexed academic publications — making it well suited to an original, self-contained descriptive study rather than a replication or extension of existing published work. Third, its structure (a mix of categorical, numerical, and short textual fields) is directly compatible with the descriptive and correlational research questions posed above, without requiring extensive feature engineering or external data linkage.

- **Source:** https://www.kaggle.com/datasets/patrasaurabh/trending-chrome-extensions
- **Number of records (extensions):** N = 50
- **Variables/columns used:** category (functional category assigned by the Chrome Web Store), rating (mean user rating, 1–5 scale), rating_count (number of user reviews underlying the rating).

  The raw file additionally includes title, url, summary, user_count, last_updated, offered_by, img_url, description, language, and developer_info, which were not used in the present analysis but are available for follow-up work (see the Future Work discussion in the Conclusion).
- **Data collection date (per source):** Not explicitly documented by the original author; inferred from the last_updated field values, which range from 2016 to 2019.
- **Missing data handling:** the language and developer_info fields contained missing values (34 and 3 rows respectively) but were not used in the RQ1–RQ3 analysis and therefore did not require imputation or row removal.

It is worth noting explicitly, as a matter of methodological transparency, that the dataset captures a cross-sectional snapshot of extensions flagged as trending by the Chrome Web Store's own (undisclosed) ranking mechanism, rather than a complete census of all extensions available in the store. It is also a comparatively small sample (N = 50), a point addressed in the Limitations.

Before any statistical analysis, the following preprocessing steps were applied to the raw CSV file:

1. Loading of the raw CSV file into a pandas DataFrame (50 rows, 13 columns).
2. Removal of duplicate entries: none were found (0 rows removed).
3. Coercion of rating and rating_count to numeric types.
4. Standardization of category labels (whitespace stripped).
5. Removal of rows with missing values in the three core analytic variables (category, rating, rating_count): none were found (0 rows removed).

The final analytic sample therefore matched the raw sample: N = 50 extensions across all analyses.

With the analytic sample established, the statistical approach follows directly from the three research questions stated above, and was chosen to be methodologically conservative: rather than fitting predictive models whose assumptions and generalizability would need extensive justification, we rely on descriptive statistics and non-parametric inferential tests appropriate for the observed distributional properties of the data.

- **Descriptive statistics (RQ1):** frequency counts and percentages of extensions per functional category; mean, median, and standard deviation of ratings and review counts.
- **Normality check:** prior to selecting correlation and group-comparison methods, we tested the normality of the rating and review count distributions using the Shapiro-Wilk test. Both variables departed significantly from normality, justifying the non-parametric methods used for RQ2 and RQ3.
- **Correlation analysis (RQ2):** given the confirmed non-normality of both variables, we used Spearman's rank correlation coefficient (ϱ) rather than Pearson's r to assess the association between review count and rating. Spearman's ϱ is appropriate here because it captures monotonic — not necessarily linear — relationships and does not assume normally distributed variables. It is computed as:

$$\rho = 1 - \frac{6\sum_i d^2}{n(n^2-1)} \quad (1)$$

where di is the difference between the ranks of the ith paired observation (review count rank minus rating rank) and n is the number of observations (n = 50).

- **Group comparison (RQ3):** to assess whether rating distributions differ across categories, we used the Kruskal-Wallis H test, the non-parametric analogue of one-way ANOVA, motivated by the same non-normality concern

and by the highly unequal category sizes in this sample (ranging from N = 1 to N = 29). The test statistic is computed as:

$$H = \frac{12}{n(n+1)} \cdot \sum_i \frac{R_i^2}{n_i} - 3(n+1) \quad (2)$$

where k is the number of groups (categories), ni is the sample size of group i, Ri is the sum of ranks in group i, and n is the total sample size across all groups (n = 50).

- **Post-hoc analysis:** a Dunn's test with Bonferroni correction was planned in the event of a statistically significant Kruskal-Wallis result; as reported in the category-comparison results below, this condition was not met, so no post-hoc comparisons were run.
- **Significance threshold:** $\alpha = 0.05$ for all inferential tests, with exact p-values reported.
- **Effect sizes:** in addition to p-values, we report epsilon-squared (ε^2) for the Kruskal-Wallis test to contextualize the practical, and not merely statistical, magnitude of the observed group differences. Epsilon-squared is computed as:

$$\varepsilon^2 = \frac{H - k + 1}{n - k} \quad (3)$$

where H is the Kruskal-Wallis statistic defined in Equation 2, k is the number of groups (categories), and n is the total sample size. This formula makes explicit why ε^2 can take a small negative value when H falls below $k - 1$ in small, unbalanced samples (as observed in the category-comparison results below): the numerator becomes negative while the denominator remains positive, an artifact conventionally reported as $\varepsilon^2 \approx 0$.

- **Confidence intervals:** to complement the point estimate, we report an approximate 95% confidence interval for Spearman's ϱ, computed via the Fisher z-transformation (appropriate given N = 50). The transformation and its inverse are:

$$z = \frac{1}{2} \ln\left[\frac{1+\rho}{1-\rho}\right], \quad SE_z = \frac{1}{\sqrt{n-3}} \quad (4)$$

$$CI_z = z \pm 1.96 \cdot SE_z \quad (5)$$

$$\rho = \frac{e^{2z} - 1}{e^{2z} + 1} \quad (6)$$

Applying this to the observed $\rho = .712\ (N = 50)$ yields the 95% CI $[.54, .83]$ reported in Table 3.

- **Software/tools:** Python 3, pandas, scipy.stats (shapiro, spearmanr, kruskal), and matplotlib/seaborn for visualization.

4. Results

We begin with the descriptive overview that addresses RQ1, before turning to the two inferential analyses that address RQ2 and RQ3.

Table 1. Distribution of extensions by category (N = 50)

Category	N	% of total
Productivity	29	58.0%
Extensions	6	12.0%
Social & Communication	5	10.0%
Shopping	4	8.0%
Fun	2	4.0%
Accessibility	2	4.0%
Themes	1	2.0%
Developer Tools	1	2.0%
Total	50	100%

The sample is heavily concentrated in the Productivity category, which accounts for more than half of all trending extensions (58.0%). The remaining seven categories are each represented by five or fewer extensions, with Themes and Developer Tools represented by only a single extension each. This distribution is far from uniform and should be kept in mind when interpreting the category-level comparison presented below, since several categories are too small ($N \leq 2$) to support robust statistical conclusions on their own.

Table 2. Descriptive statistics and normality test results for rating and review count (N = 50)

Variable	Mean	Median	SD	Min	Max	Shapiro-Wilk W	p
Rating	3.84	4.00	0.83	1.40	4.90	0.908	< .001

Variable	Mean	Median	SD	Min	Max	Shapiro-Wilk W	p
Review count (rating_count)	33,367.26	5,275.50	71,216.35	235	350,370	0.507	< .001

Ratings are moderately high on average (M = 3.84 out of 5) with a median of exactly 4.0, and a standard deviation (0.83) indicating meaningful but not extreme spread — the bulk of extensions cluster between roughly 3.5 and 4.5, with a minority of low outliers (minimum rating of 1.4). Review counts, by contrast, show a very large gap between the mean (33,367) and the median (5,276), with a standard deviation (71,216) more than double the mean — a strong signature of right-skew consistent with a small number of extensions accumulating a disproportionate share of total reviews, as is typical of engagement metrics in digital marketplaces. Both variables also departed significantly from a normal distribution (Shapiro-Wilk $p < .001$ for each), which motivates the non-parametric approach used for RQ2 and RQ3.

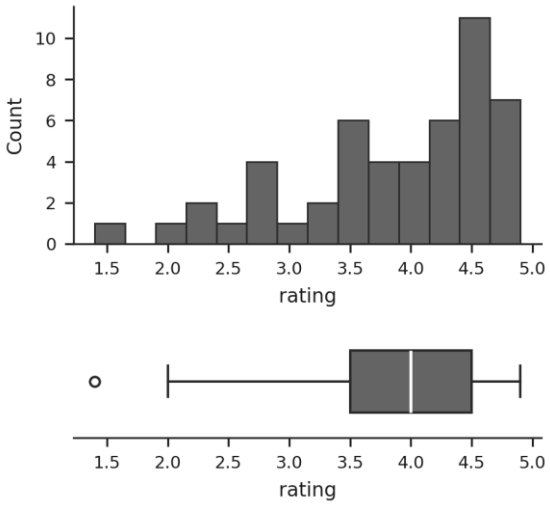

Figure 1. Histogram and boxplot of the rating distribution.

The right-skew of review counts described above is directly visible in Figure 2: the bulk of extensions cluster in the low thousands of reviews, with a small number of high-visibility extensions extending the distribution out to the hundreds of thousands (log scale), consistent with the markedly lower Shapiro-Wilk W (0.507) reported for this variable in Table 2.

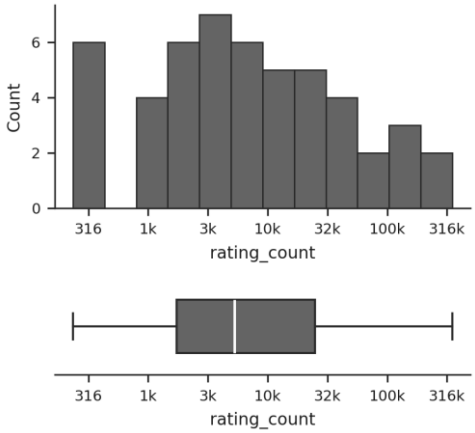

Figure 2. Histogram and boxplot of the review count distribution (log scale).

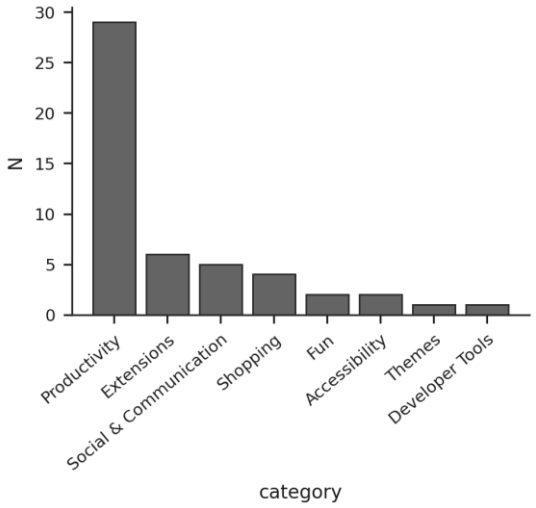

Figure 3. Number of extensions per category.

Table 3 summarizes the two inferential tests conducted to address RQ2 and RQ3, each discussed in detail below.

Table 3. Summary of inferential tests for RQ2 and RQ3 (N = 50)

Test	RQ	Statistic	p	95% CI / Effect size	Interpretation
Spearman's rank correlation (ϱ)	RQ2	ϱ = .712	< .001	95% CI [.54, .83]	Strong, positive, statistically significant
Kruskal-Wallis H	RQ3	H(7) = 5.21	.635	$\varepsilon^2 = -0.043$ (≈ 0)	Not statistically significant; negligible effect

Note. The 95% confidence interval for Spearman's ϱ was computed via the Fisher z-transformation (N = 50). ε^2 denotes epsilon-squared, the effect size associated with the Kruskal-Wallis test.

Having established the descriptive baseline for RQ1, we turn next to the correlation analysis that addresses RQ2.
Spearman's correlation between review count and rating: $\rho = .712, p < .001\ (p = 6.79 \times 10^{-9}), 95\%\ \text{CI}\ [.54, .83]$ (Fisher z-transformation).
This constitutes a strong, positive, and highly statistically significant monotonic association: extensions with more reviews in this sample also tended to have higher average ratings (see Table 3 for a consolidated summary).
As a robustness check, Table 4 compares this result against Pearson's linear correlation coefficient, computed both on the raw (untransformed) variables and on log-transformed review count.
The substantial gap between Pearson's r on raw data and Spearman's ϱ confirms, quantitatively, that the non-parametric approach used here was not merely a formal precaution: a linear correlation on the untransformed data would have understated the strength of the review count–rating association by nearly half, an artifact of the extreme right-skew and high-leverage outliers documented in Table 2 and Figure 2.

Table 4. Robustness check: correlation coefficient by method (N = 50)

Method	Statistic	p
Pearson's r (raw variables)	r = .397	.004
Pearson's r (log-transformed review count)	r = .648	< .001
Spearman's ϱ (reported result)	ϱ = .712	< .001

Note. Pearson's r on raw variables is attenuated by the extreme right-skew of review count (Table 2, Figure 2); the log transformation partially corrects for this, but Spearman's ϱ remains the appropriate and most robust estimate given the confirmed non-normality of both variables discussed above.

A second, independent robustness concern is compositional rather than statistical: because Productivity alone accounts for 58% of the sample (Table 1), the overall RQ2 correlation could in principle be an artifact of between-category composition rather than a genuine within-category association — a pattern sometimes called Simpson's paradox.
To rule this out, we recomputed Spearman's ϱ separately within Productivity and within the seven remaining categories combined (Table 5).

Table 5. Robustness check: Spearman's ϱ by category composition (N = 50)

Subset	N	ϱ	p	Interpretation
Overall (reported result)	50	ϱ = .712	< .001	Strong, statistically significant (Table 3)
Productivity only	29	ϱ = .825	< .001	Even stronger within the dominant category
Excluding Productivity	21	ϱ = .373	.096	Weaker; not statistically significant at this N

Note. "Overall" reproduces the reported RQ2 result from Table 3. "Productivity only" and "Excluding Productivity" recompute Spearman's ϱ within each subset of the same N = 50 sample.

The association is, if anything, stronger within Productivity alone ($\rho = .825$) than in the full sample, and it is Productivity — not the smaller, more heterogeneous remainder of the sample — that drives the overall result.
This pattern is inconsistent with a Simpson's-paradox-style explanation in which the marginal correlation is manufactured by between-category composition; instead, it indicates that the review count–rating association is a genuine within-category phenomenon, at least within the dominant category. The weaker, non-significant estimate outside Productivity ($\rho = .373, p = .096$) should be read in light of its sharply reduced statistical power (N = 21, spread thinly across seven categories) rather than as evidence of a qualitatively different relationship; we return to this point in the Limitations.

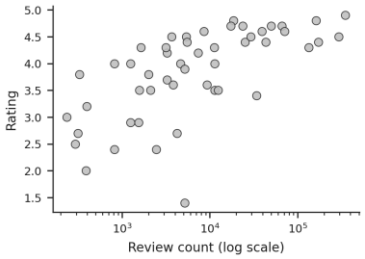

Figure 4. Scatter plot of review count (log scale) versus rating.

Figure 5 makes this compositional check visually explicit by distinguishing Productivity extensions from the rest of the sample.
The positive review count–rating trend visibly holds within the Productivity subgroup on its own — including at the high end of the review-count range — rather than being carried disproportionately by the smaller, more scattered set of non-Productivity extensions.

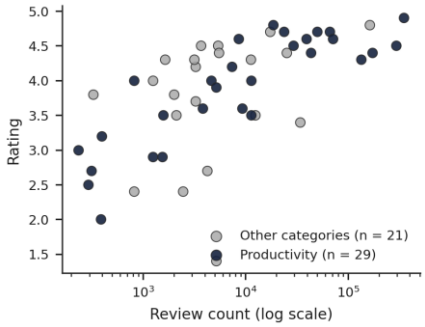

Figure 5. Scatter plot of review count (log scale) versus rating, by Productivity vs. other categories.

To ground this aggregate result in concrete terms, Table 6 lists the eight individual extensions with the highest review counts in the sample.

These are, unsurprisingly, well-known, high-utility tools — ad blockers and VPN proxies chief among them — and all eight carry ratings at or above 4.3, consistent with the strong review count–rating association reported above holding not only in aggregate but at the level of individually recognizable products.

The concentration of Productivity-category extensions at the top of this list (seven of eight) also illustrates directly, rather than only statistically, why that category drives the overall RQ2 correlation (Table 5).

Table 6. Most-reviewed extensions in the sample (top 8 by review count, N = 50)

Extension	Category	Rating	Review count
Hola Free VPN Proxy Unblocker – Best VPN	Productivity	4.9	350,370
AdBlock — best ad blocker	Productivity	4.5	295,307
Adblock Plus – free ad blocker	Productivity	4.4	172,368
Honey	Shopping	4.8	162,447
Adblock for Youtube™	Productivity	4.3	134,018
Touch VPN – Secure and unlimited VPN proxy	Productivity	4.6	70,847
Tampermonkey	Productivity	4.7	66,108
AdGuard AdBlocker	Productivity	4.7	50,049

Note. Ratings and review counts as recorded in the source dataset (Patra, 2019); titles reproduced verbatim from the Chrome Web Store listing.

Turning to RQ3, the Kruskal-Wallis test result: $H(7) = 5.21, p = .635$ (Table 3). This result is not statistically significant at the $\alpha = 0.05$ threshold.

Effect size: $\varepsilon^2 = -0.043$, interpreted as an essentially null effect ($\varepsilon^2 \approx 0$); see Methods for why the formula can yield a slightly negative value in small, unbalanced samples such as this one.

Table 7. Mean and median rating by category, ordered by median rating (descending)

Category	N	Mean rating	Median rating	SD
Shopping	4	4.28	4.30	0.45
Themes	1	4.20	4.20	—
Extensions	6	3.85	4.15	0.80
Accessibility	2	4.00	4.00	0.71

Category	N	Mean rating	Median rating	SD
Productivity	29	3.90	4.00	0.79
Social & Communication	5	3.56	3.80	0.74
Fun	2	3.70	3.70	1.41
Developer Tools	1	1.40	1.40	—

Given the non-significant Kruskal-Wallis result, no post-hoc pairwise comparisons (e.g., Dunn's test) were conducted, consistent with the analysis plan in Methods.

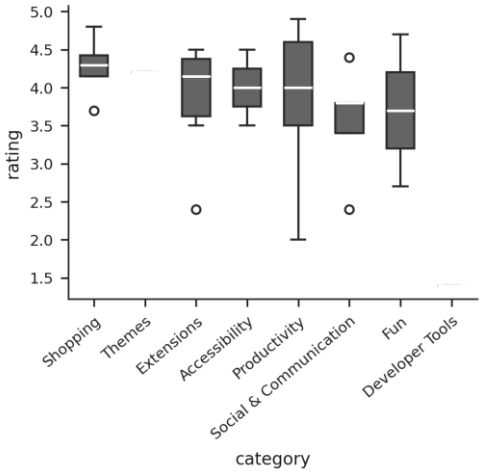

Figure 6. Boxplot of rating by category, ordered by median.

5. Discussion

The results paint a fairly coherent picture. The compositional finding for RQ1 — a strong concentration of trending extensions in the Productivity category — suggests that whatever mechanism the Chrome Web Store uses to surface "trending" extensions, it disproportionately favors utilitarian, task-oriented tools over entertainment, social, or thematic extensions.

This is a purely descriptive observation, but it usefully contextualizes the other two findings: with 58% of the sample concentrated in one category, statistical power to detect between-category differences (RQ3) is inherently limited by design, not only by sample size.

The RQ2 finding — a strong positive correlation between review count and rating ($\rho = .71, p < .001$) — indicates that, in this sample, extensions that accumulated more user reviews were also, on average, rated more favorably.

This is consistent with a self-reinforcing dynamic in which well-liked extensions attract more installations and thus more reviews, which in turn increases their visibility within the store's trending mechanism — though this correlational design cannot establish which direction (if either) is causal, nor rule out shared underlying causes (e.g., overall product quality driving both). Nor is it an artifact of the sample's compositional skew toward Productivity: the association held — if anything, more strongly — within that category alone (Table 5), so category dominance does not appear to be manufacturing the marginal result.

The RQ3 finding — no statistically significant difference in ratings across categories — suggests that, once an extension is popular enough to be "trending," its functional category is not a strong independent predictor of how favorably it will be rated. Put differently, user satisfaction (as proxied by rating) appears to be more a property of individual extensions than of the categories they belong to. This null result should be interpreted cautiously, however (see Limitations).

Set against the related work reviewed earlier, the strength of the review count–rating correlation observed here ($\rho = .71, p < .001$) is broadly consistent with, if somewhat stronger than, the positive rating–popularity associations reported by Finkelstein et al. (2017) in the BlackBerry World App Store, lending external plausibility to the direction and general magnitude of the association found in the present, much smaller sample. The concentration of trending extensions in a small number of categories similarly echoes Zhong and Michahelles's (2013) "superstar market" pattern, suggesting that this structural feature may generalize across different types of digital software marketplaces rather than being specific to mobile apps.

The RQ3 null result (no significant rating difference across categories) is more difficult to reconcile with prior work. Tian et al. (2015) found that app category was one of 17 (of 28) factors significantly associated with high- versus low-rated status among Android apps, which might lead one to expect a similar category effect here. Two non-mutually-exclusive explanations seem plausible. The most straightforward is statistical power: the present study's power was far lower than Tian et al.'s large-scale comparison, so a true but modest category effect could easily go undetected here (see Limitations).

A second explanation draws on Hu, Bezemer, and Hassan (2018), who showed, in a study of star ratings for cross-platform Android/iOS apps, that ratings are shaped by factors extending well beyond an app's category or objective quality — including platform-specific user expectations and complaint patterns that do not map cleanly onto category boundaries. If rating is similarly driven in the Chrome Web Store by extension-specific factors (execution quality, responsiveness to user feedback, permission scope) more than by category-level norms, a weak or null category effect would not be surprising even with larger samples. Distinguishing between these two explanations is not possible with the present dataset and is flagged as a direction for future work in the Conclusion.

These findings also carry some practical implications. The strong positive association between review volume and rating (RQ2) suggests that, at least descriptively, review count could serve as a weak proxy signal for quality in this marketplace — though given the correlational and cross-sectional nature of this study, this should not be read as a recommendation to rely on review count alone for quality assessment.

The null result for RQ3 tentatively suggests that store curators or developers should not assume that certain categories are inherently disadvantaged in terms of achievable ratings; individual execution may matter more than category positioning. Both implications, however, are offered cautiously given the sample size, and the following limitations should be kept in mind when interpreting all three findings.

- The dataset represents a snapshot of trending extensions at a single point in time and may not generalize to the full population of extensions available in the Chrome Web Store, many of which are neither trending nor widely reviewed.
- Sample size: with only $N = 50$ extensions spread across eight categories — several with $N \leq 2$ — the Kruskal-Wallis test for RQ3 had limited statistical power to detect anything but very large between-category differences. The non-significant result should therefore be interpreted as "no evidence of a difference detectable at this sample size," not as strong evidence of true equivalence across categories. The same caution applies to the non-Productivity subgroup in the Table 5 robustness check ($N = 21$): its weaker, non-significant ϱ is consistent with reduced power at that sample size and should not be read as evidence that the review count–rating association is absent outside Productivity.
- The correlational, cross-sectional design means the strong RQ2 association ($\varrho = .71$) cannot establish causal direction; it is equally consistent with popularity driving favorable perception, with underlying product quality driving both engagement and satisfaction, or with the store's own ranking algorithm feeding back on itself by surfacing already well-rated extensions to more users.
- The dataset does not include variables such as true installation counts (beyond the coarse user_count field not used here), extension age since publication, or permission scope, all of which could plausibly confound or explain the associations reported here.
- Ratings and review counts on the Chrome Web Store are subject to well-documented platform-level biases (e.g., review solicitation prompts, potential for review manipulation) that are not observable or correctable within the scope of this dataset.
- The categorical labels used in the dataset reflect the Chrome Web Store's own taxonomy (e.g., a broad "Extensions" category alongside more specific ones like "Shopping"), which may group functionally dissimilar extensions together or split similar ones apart in ways that are not fully transparent to external analysts.

6. Conclusion

This study presented a descriptive and correlational analysis of 50 trending Google Chrome extensions, examining how they are distributed across functional categories and how review volume and category relate to user ratings.

Extensions were found to be heavily concentrated in the Productivity category, review count was strongly and positively correlated with rating, and no statistically significant differences in rating were found across categories.

These findings offer an initial, transparently documented empirical baseline for a browser extension marketplace that has received little prior analytical attention, and they underscore the value of establishing simple descriptive results before pursuing more complex predictive or causal modeling on similar data. Building on this baseline, several directions for future work follow naturally:

- Extension of this analysis to a larger and/or longitudinal dataset tracking the same extensions over time, which would both increase statistical power for category-level comparisons and allow the review–rating relationship to be examined dynamically.
- Incorporation of text-based features (e.g., sentiment or keyword themes from the summary and description fields available in the raw dataset but not used here) as a follow-up study.
- Incorporation of the user_count field, which was present in the raw data but excluded from the present analysis, to examine whether installation counts track review counts and ratings in the same way.
- Comparison across multiple browser extension marketplaces (Chrome, Firefox, Edge) to assess whether the patterns documented here are specific to the Chrome Web Store or reflect more general dynamics of browser extension marketplaces.

Data Availability Statement

The dataset used in this study is publicly available at https://www.kaggle.com/datasets/patrasaurabh/trending-chrome-extensions, released under a CC0: Public Domain license. The analysis code is not being made available in a public repository at this time.

Conflicts of Interest

The author(s) declare no conflict of interest.

References

1. Patra, S. (2019). Trending Google Chrome Extensions [Data set]. Kaggle. https://www.kaggle.com/datasets/patrasaurabh/trending-chrome-extensions
2. Tian, Y., Nagappan, M., Lo, D., & Hassan, A. E. (2015). What are the characteristics of high-rated apps? A case study on free Android applications. In 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 301–310). IEEE. https://doi.org/10.1109/ICSM.2015.7332476
3. Potharaju, R., Newell, A., Nita-Rotaru, C., & Zhang, X. (2018). A Longitudinal Study of Google Play. arXiv:1802.02996 / IEEE. https://arxiv.org/abs/1802.02996
4. Carlini, N., Felt, A. P., & Wagner, D. (2012). An evaluation of the Google Chrome extension security architecture. In Proceedings of the 21st USENIX Security Symposium (pp. 97–111). USENIX Association. https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/carlini
5. Finkelstein, A., Harman, M., Jia, Y., Martin, W., Sarro, F., & Zhang, Y. (2017). Investigating the relationship between price, rating, and popularity in the BlackBerry World App Store. Information and Software Technology, 87, 119–139. https://doi.org/10.1016/j.infsof.2017.03.002
6. Kitchenham, B., Madeyski, L., Budgen, D., Keung, J., Brereton, P., Charters, S., Gibbs, S., & Pohthong, A. (2016). Robust statistical methods for empirical software engineering. Empirical Software Engineering, 22(2), 579–630. https://doi.org/10.1007/s10664-016-9437-5
7. Zhong, N., & Michahelles, F. (2013). Google Play is not a long tail market: An empirical analysis of app adoption on the Google Play app market. In Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC '13) (pp. 499–504). ACM. https://doi.org/10.1145/2480362.2480460
8. Hu, H., Bezemer, C.-P., & Hassan, A. E. (2018). Studying the consistency of star ratings and the complaints in 1 & 2-star user reviews for top free cross-platform Android and iOS apps. Empirical Software Engineering, 23(6), 3442–3475. https://doi.org/10.1007/s10664-018-9604-y
9. Aggarwal, A., Viswanath, B., Zhang, L., Kumar, S., Shah, A., & Kumaraguru, P. (2018). I spy with my little eye: Analysis and detection of spying browser extensions. In 2018 IEEE European Symposium on Security and Privacy (EuroS&P) (pp. 47–61). IEEE. https://doi.org/10.1109/EuroSP.2018.00012