MOZILLA FIREFOX ANALYSIS PROJECT

By Daniel Clark

AGENDA

- The Ask
- The Data
- The Subjects
- The Limitations
- The Assumptions
- The Recommendation
- The Proof
- The Next Steps
- The Questions
- The Appendix

THE ASK

Use PGAdmin databases developed by Mozilla's engineers to substantiate a redesign of the Firefox bookmarks feature in the upcoming version update.

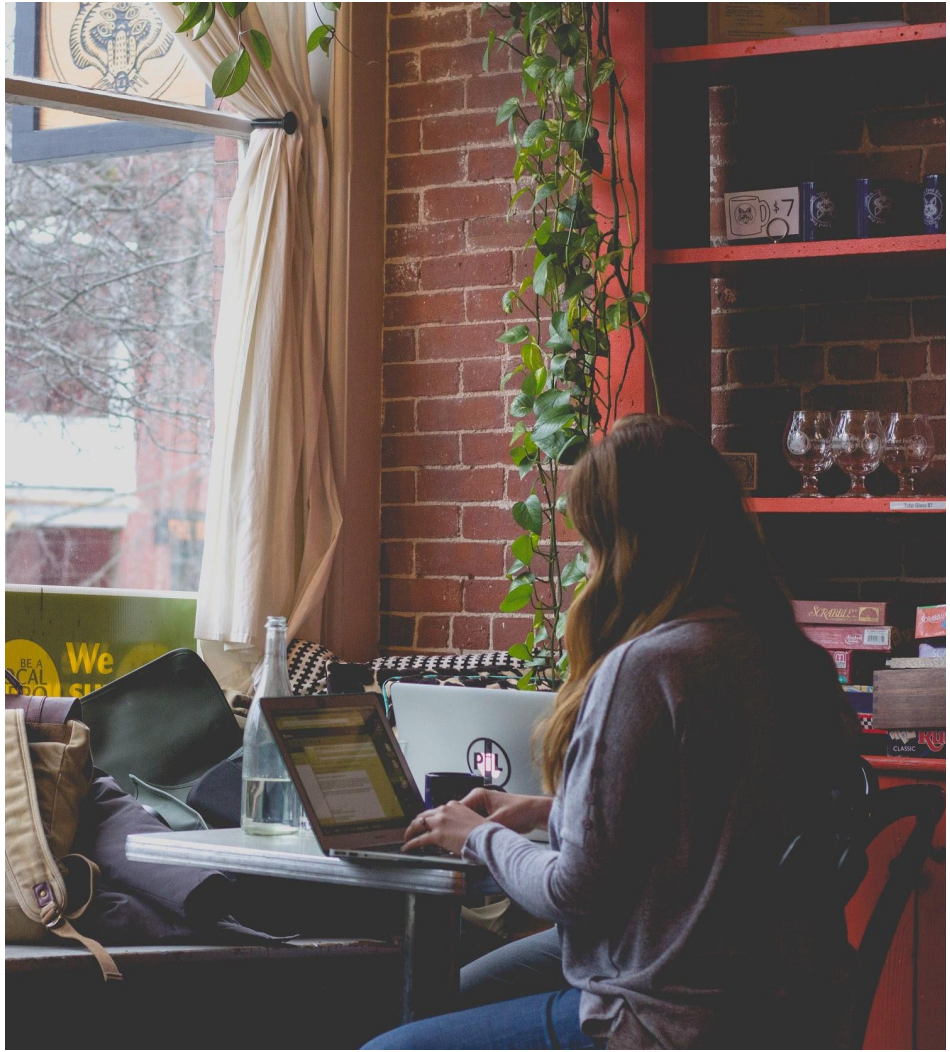
My very opinionated colleague thinks that tabs are more interesting and that we should focus on optimizing those.

Which direction should we go?

THE DATA

2016 Outbrain event data from 6/10/2010

- Event data logging all the actions from participants in their browsing sessions on Firefox
- User Data outlining the OS and Firefox version used
- Sample surveys of participant demos and browsing behaviors
- All tables have user id data available, which can be used to draw correlations between multiple tables



THE SUBJECTS

- 4,042 people completed the survey
- 14,718 distinct users on the event table
- 27,267 distinct users in the User ID table

Nearly 2/3rds of respondents are under 25

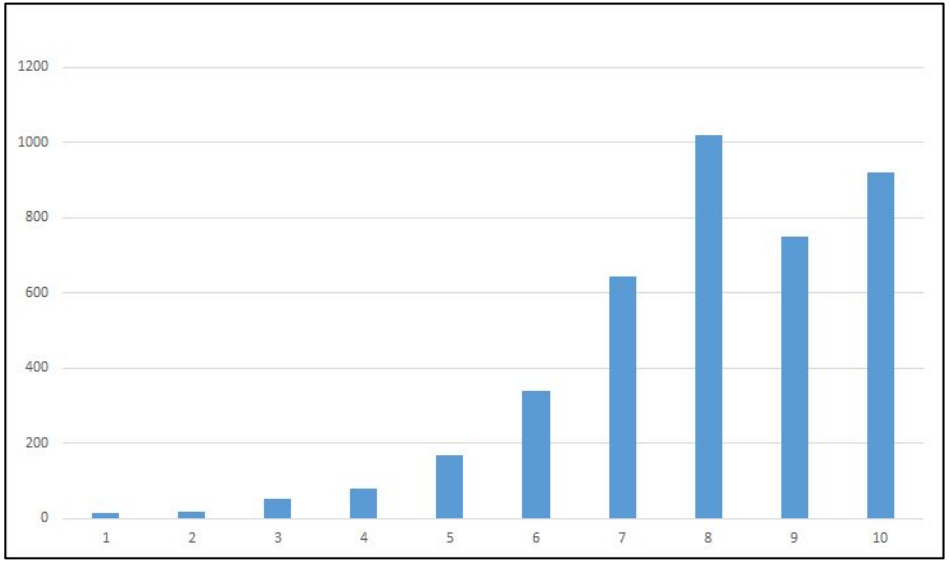
More than 90% of respondents are male

THE LIMITATIONS

Does this data represent all Firefox users?

No!

- Predominantly Male
- Predominantly Young
- Data skews more towards the more skilled computer users



Self-proclaimed computer Skill level (1 being caveman - 10 being Bill Gates)

THE ASSUMPTIONS

Does this mean that our data is useless?

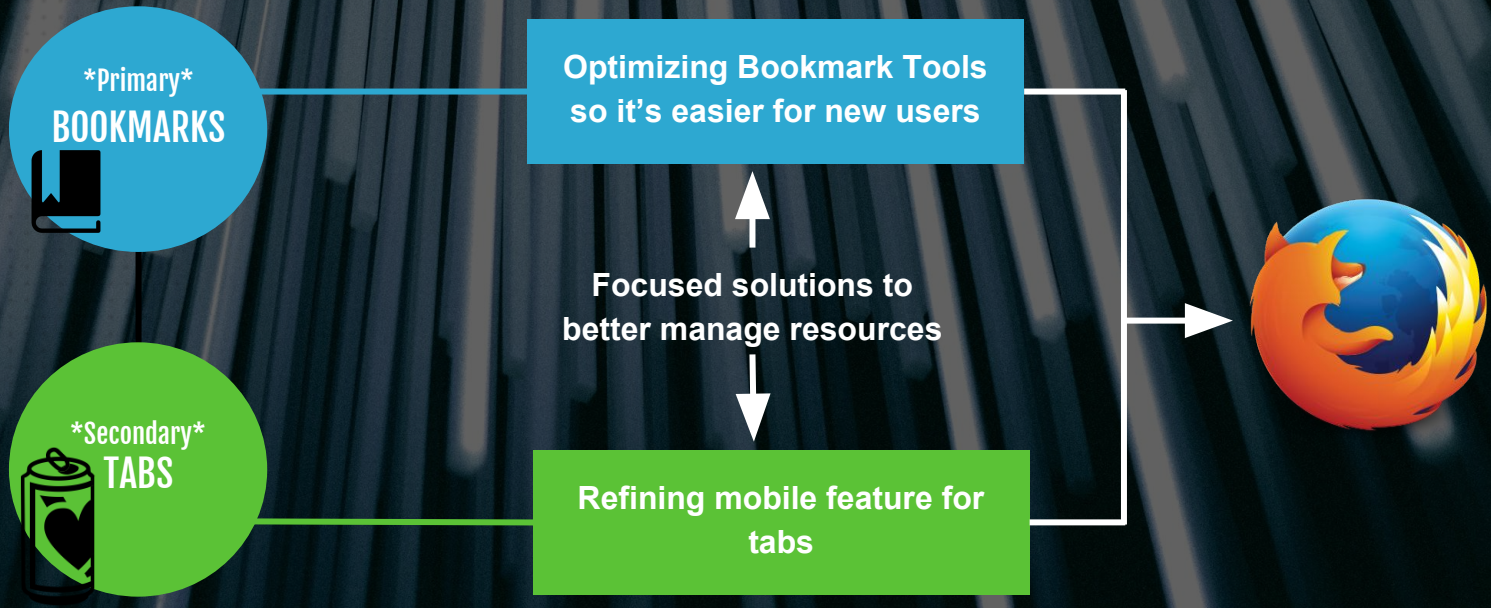
Not Necessarily!!

This means that we have to prioritize our decision into the following 3 criteria:

- Focused Optimization
- Resource Allocation
- Best for Business

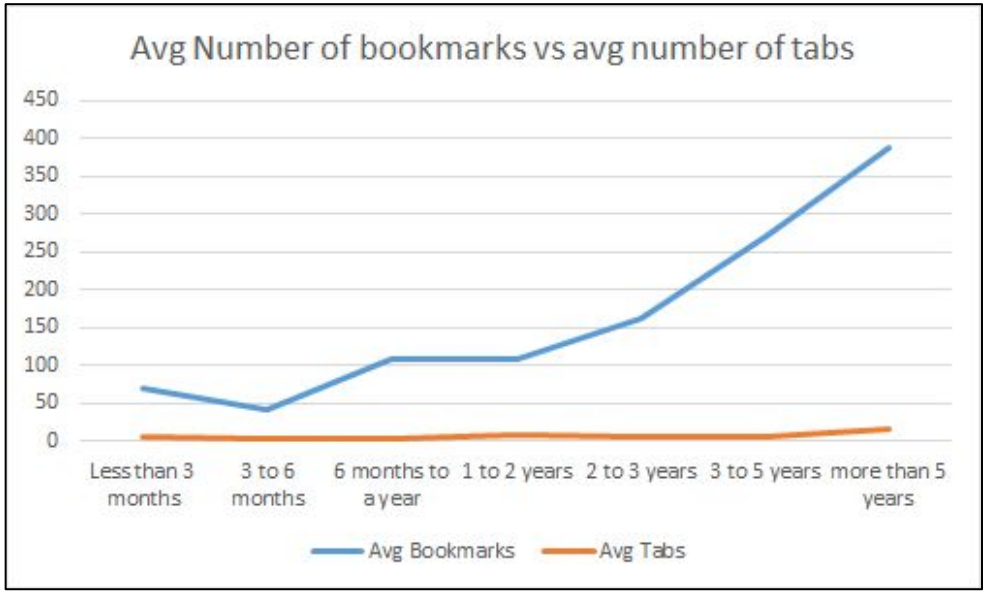
THE RECOMMENDATION

Focus on improving specific aspects of the tabs and bookmarks rather than needlessly spend resources overhauling the entire feature.



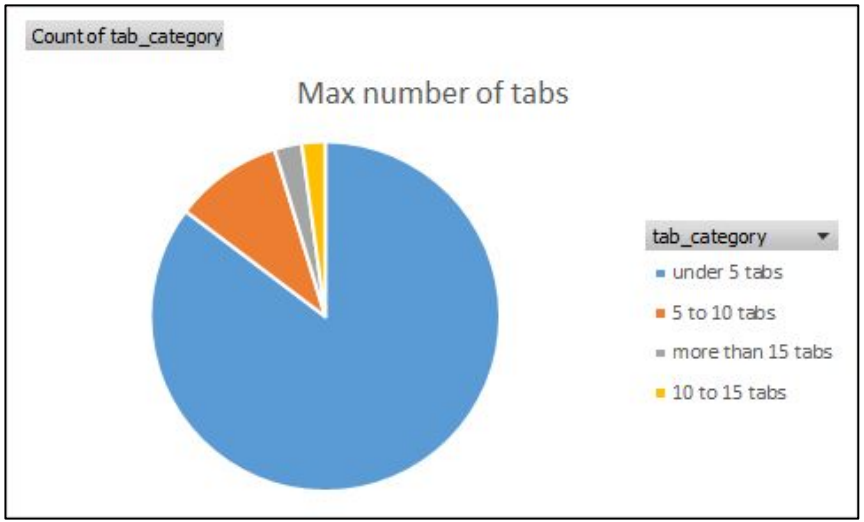
THE PROOF

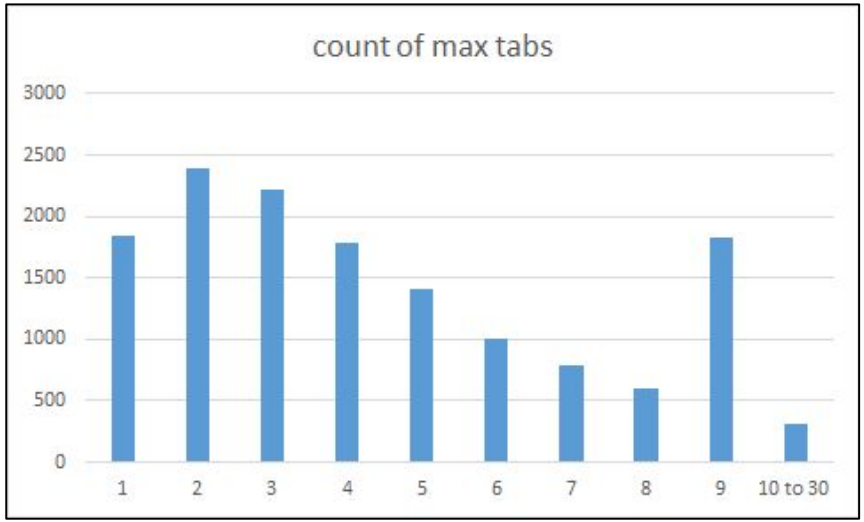
- Looking at the data on the average number of tabs that each user maxed out on, as well as the average number of bookmarks, we can see that the number of bookmarks used rises much faster as users get more experience with Firefox.
- **With that said, only 24% of users launched a bookmark during sample week.**



The average number of bookmarks increase at a much faster rate than the average number of tabs as people get more experience with Firefox

THE PROOF





As you can see here, the majority of users primarily max out at under 10 tabs with a few outliers going above.

THE PROOF

8.8

Avg. Number of Tabs for Non-Mobile Users

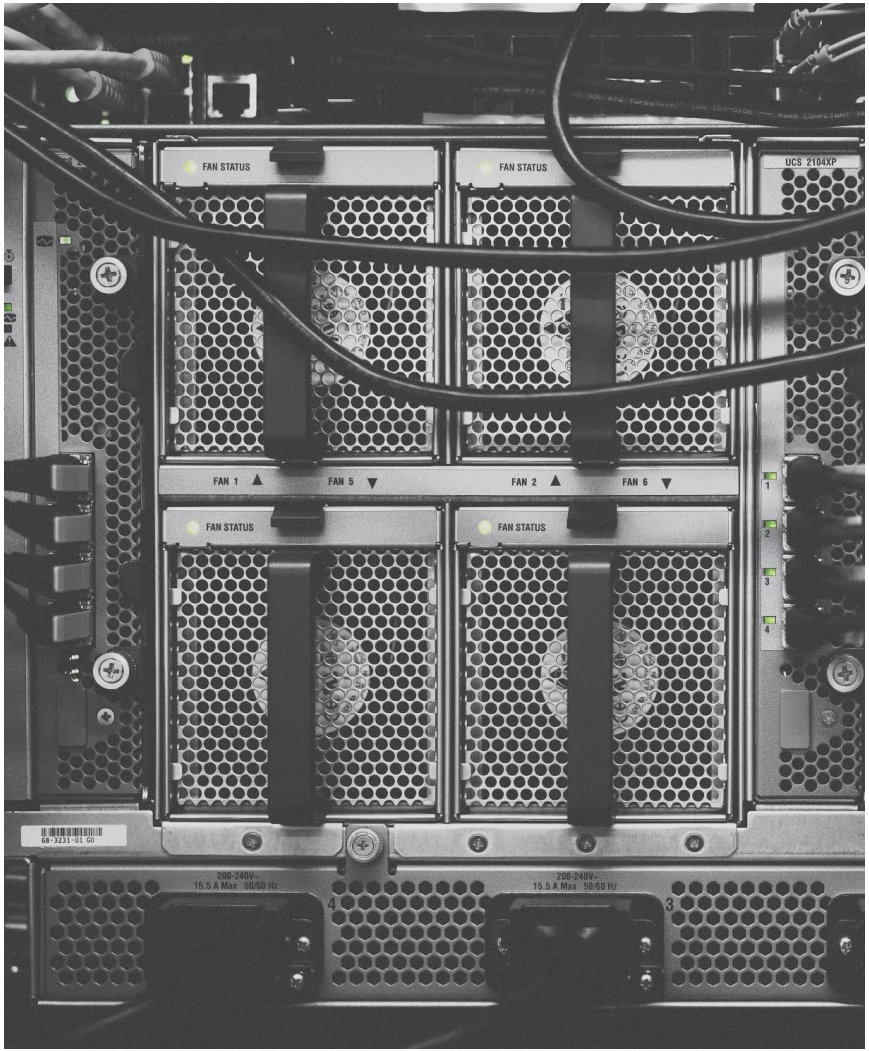
10.3

Avg. Number of Tabs per User

12.9

Avg. Number of tabs on Mobile

Tabs accumulate much more among mobile users than non-mobile users.

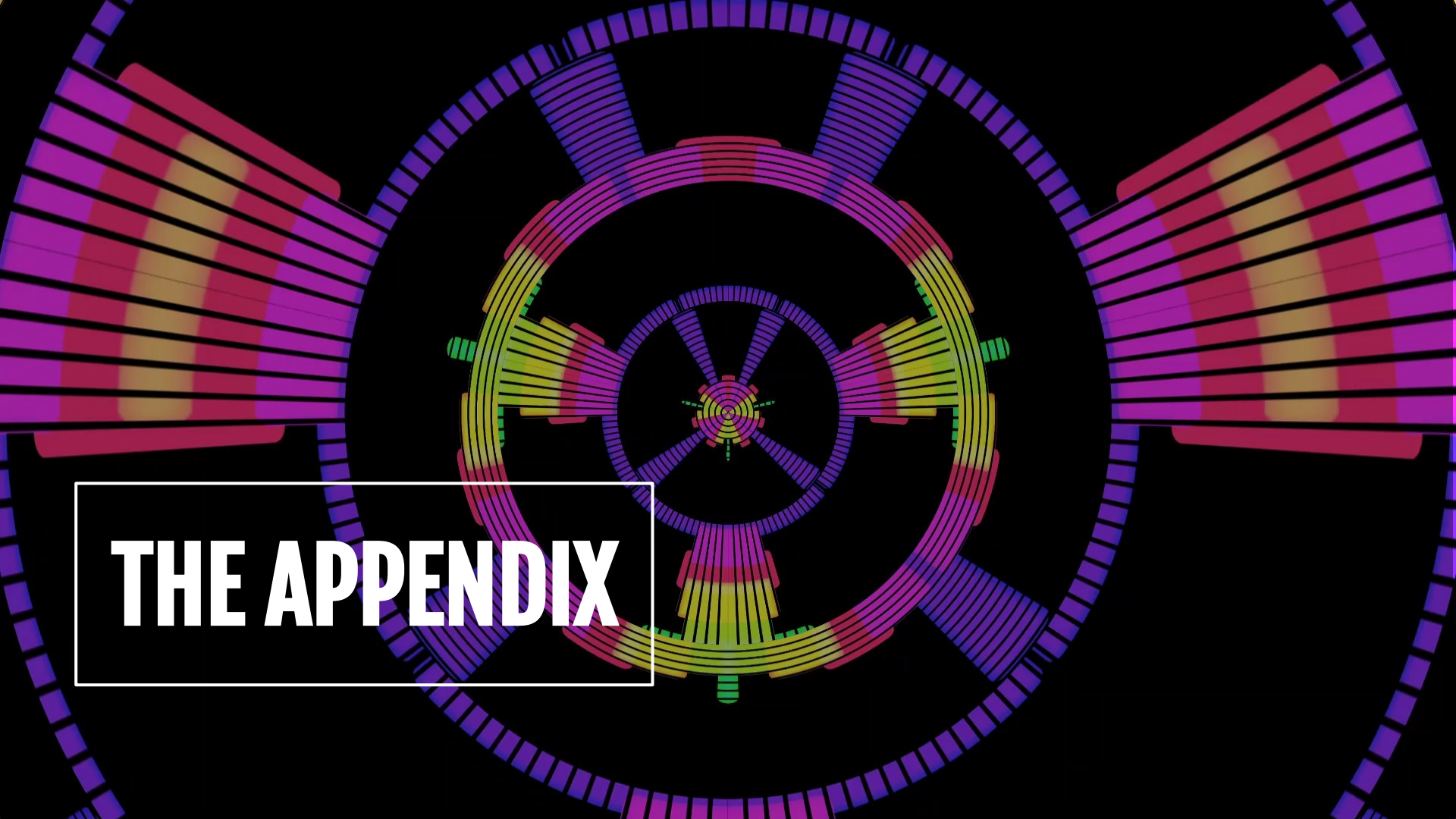


THE NEXT STEPS

- Focus primarily on optimizing the Bookmark feature with emphasis on the following audiences:
  - Users on Firefox for less than a year
- Consider optimizing the tabs feature for Mobile Users in future version updates.
- For future research endeavors:
  - Consider revisiting survey methodology to capture a wider span of users with different skill levels with the internet.
  - Track the age of bookmarks and the amount of times that bookmark is referred to during a session.
  - **Capture more data from mobile usage.**

THE QUESTIONS

Thank You


THE APPENDIX

THE CHARACTERS

- **Summarize the data in your presentation:**

- This data includes information on the user's, survey responses, and the events that took place each time the user logged on to the Firefox browsing session. This data will inform your thinking on a question about bookmarks and large quantities of tabs for the next firefox update or extension.

- **Review the Users table. Summarize the users represented in the survey.**

- select * from users join survey on users.id = survey.user_id limit 100
- Joining the Users and Survey table will provide a complete list of all the user profiles available in the database as well as some comprehensive survey data on their internet habits and skill levels across each. This is a good starting point for the analysis to get a sense of the people who represent the data and their understanding of the nuances of the interwebs.

- **Review the Survey table. How many of the total users completed the survey?**
  - select distinct(user_id), * from survey where q9 is null
  - Since there are 39 null responses in the 4081 person survey, you can do the math and say there are 4042 completed responses to the survey.

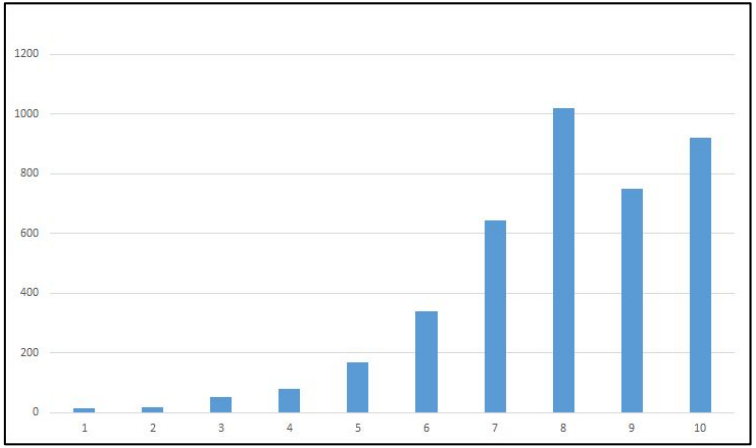
- **Of users who completed the survey, identify the number of users who are new to Firefox, as well as those who are longtime users.**
  - [note: run a case to find people who answer 0,1,2 and identify them as "new" vs people who answer 3, 4, 5, 6]
  - select count(*), case when cast(survey.q1 as int) < 3 then 'New_user' else 'seasoned_vet' end as experience from survey group by survey.q1
  - based on the query, I found that 351 respondents have been using Firefox for less than a year and 3,725 have been using for over a year. So the vast majority of the survey respondents are veterans of Firefox

THE PROBLEM

- I was asked to use the PGAdmin data to help substantiate areas of focus for the next version of Firefox. I want to revamp the bookmarks experience, but she wants me to help her substantiate her request to improve support for a large number of tabs. To find this,

PLOT / EVENTS

- **Do i think I have sufficient data to suggest that my data represents all firefox users**
- yes in the sense that just about every distinct user id (select distinct (e.user_id),* from survey e join users u on e.user_id = u.id) has data on the survey, events and users table. I think there would be a problem if we had a large number of user ids on the events table that did not appear in the users and survey tables
- However, looking at our survey data on self-attributed "computer skill level", the data shows that our users tend to skew more towards skilled computer users. So Sample Bias may be apparent.

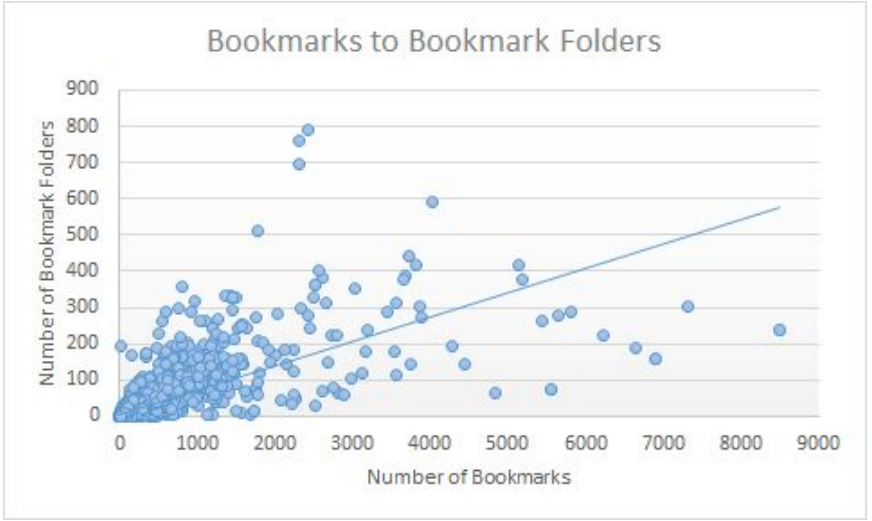


- **Key points to support focus on tabs/bookmarks**
  a. tabs - people tend to have a lot of tabs when they are browsing the internet, with no way to organize or prioritize. people have a narrow number of websites they browse
  b. Bookmarks - potentially have an account that can be accessed across different locations where people browse the internet. a strong bookmark feature can go across accounts. people have a wide variety of websites that they browse

- **Describe any initial findings related to your research questions.**
  - select * from events where event_code = 26 (Just wanted to isolate the num_tabs event code cases and then export to excel to see what the average number of windows and tabs people have open at a given time. Answer = 1.19 Windows and 6.09 Tabs.
  - select * from survey (since there are 4000 responses to the survey data, I just pulled it into an excel grid to get a sense of what people are doing on the internet. First, I wanted to see the which platforms are the most popular when browsing the internet and saw that 45% of users browse at home, 27% at school, and 17% at school and 11% mobile. I then checked to see what are the most common reasons to use the internet, I found that it was primarily socializing, communication and personal life assistance. this was followed by school and work.

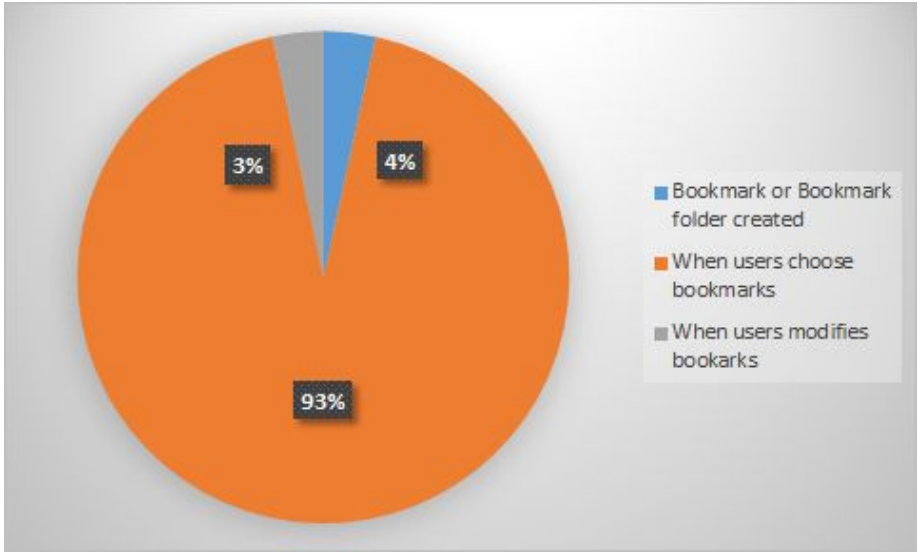
What's the median number of bookmarks? What's the average number?

- Select user_id, data1, data2, data3 from events where event_code = 8 (also: Select distinct user_id, data1, data2, data3 from events where event_code = 8)
- I made separate excel spreadsheets pulling specifically just the bookmark data as well as the distinct user bookmark data. I checked the media total bookmarks and found 28(29 distinct) bookmarks as the median
- I also looked into the average total bookmarks from the users and saw that there were 106.7 bookmarks on average amongst the total usings during the sample week. Among distinct users, total bookmarks were 123.15. So this looks like the average firefox user uses a lot of bookmarks.
- I also did an analysis checking into a scatter plot of the amount of bookmarks to bookmark folders (while moving outliers of over 1000 bookmark folders and 10,000 bookmarks. After setting a trend line, it appears that people open one bookmark folder per 15 bookmarks.



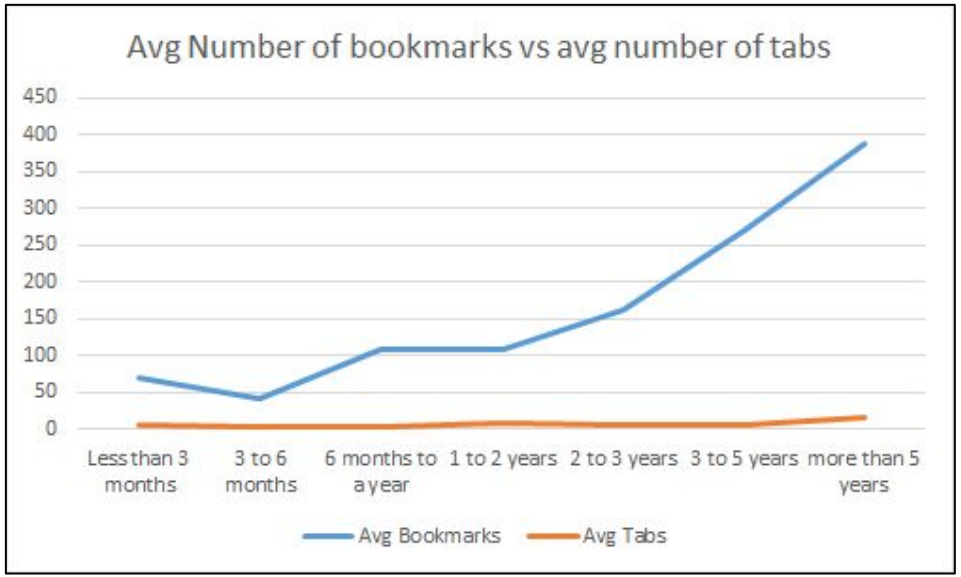
What's the distribution of how often bookmarks are used?

- Assume that you can simply count each event code involving use of a bookmark(rather than registering bookmark numbers), here we can make a chart that outlines what each count entails
- Select count(event_code) from events where event_code = 11 or event_code = 10 or event_code = 9 (take the counts of each and put into a chart)
- Looking at the chart below, 93% of actions taken during the survey week involving bookmarks related to choosing bookmarks whenever users are browsing. This is far greater than the amount of times users modify and create bookmarks.



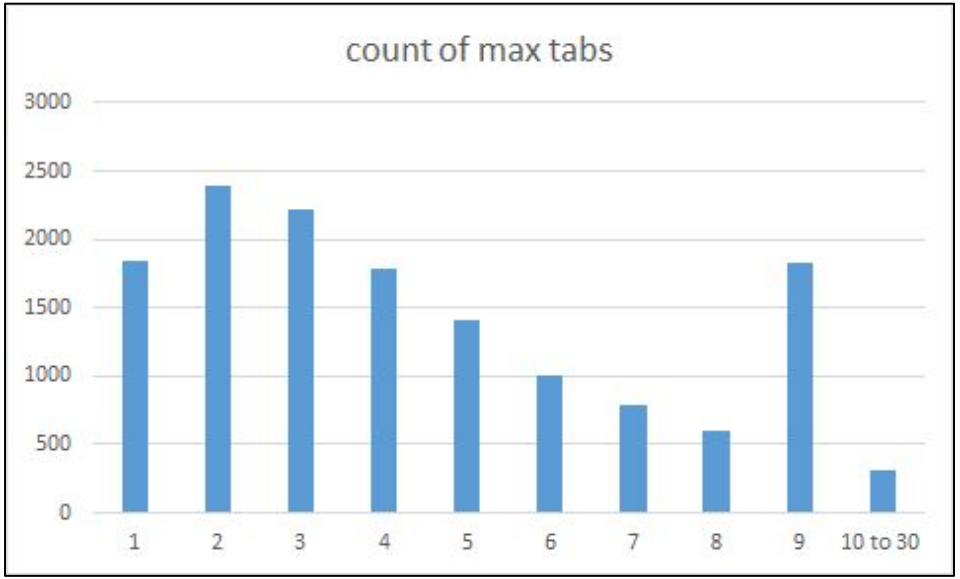
- **How does number of bookmarks correlate with how long the user has been using Firefox?**

- select distinct user_id, q1, data1 from survey inner join events using (user_id) where event_code = 8
- Looking at the specific tab data (select distinct user_id, q1, data1, data2 from survey inner join events using (user_id) where event_code = 26), The distribution for tab data is just about as balanced as the progression of bookmark data with experience with mozilla firefox. one thing that led me curious is why the average number of bookmarks per user droops at the 3 month and 2 year mark.



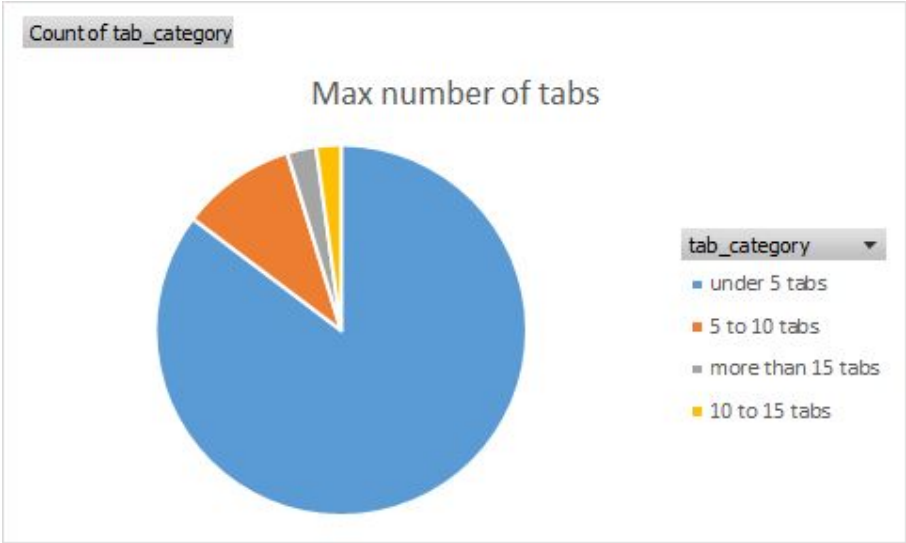
What is the distribution for the max number of tabs

- select events.user_id, max(data2) from events where event_code = 26 group by user_id
- The data query was designed to show the maximum number of tabs shown for each user_id. After doing a count, we saw that the mass majority of the distribution was dispersed pretty heavily under 10 with a few outliers over 10 at a time.



- **Are there users who regularly have more than 10 tabs open?**
    a. select user_id, avg(rtrim) as many_tabs from (select user_id, cast(trim(trailing 'tabs' from data2) as int) from events where event_code = 26) as temp group by user_id order by many_tabs desc
    b. this query is designed to designate the number of tabs column as an integer and then calculate the average occurrences under each user. After sorting by descending, I determined that there are 670 users who regularly have more than 10 tabs open. Out of the 27,337 distinct users in the events database, this makes up 2.5% of the users.

- **What fraction of user have ever had more than five tabs open? What fraction of users have ever had more than 10 tabs open? What fraction of users have had more than 15 tabs open?**
    - Select *, case when many_tabs < 5 then 'under 5 tabs' when many_tabs < 10 then '5 to 10 tabs' when many_tabs < 15 then '10 to 15 tabs' else 'more than 15 tabs' end as tab_category from (select user_id, avg(rtrim) as many_tabs from (select user_id, cast(trim(trailing 'tabs' from data2) as int) from events where event_code = 26) as temp group by user_id order by many_tabs desc) as temp
    - This function was designed to pull the maximum number of tabs of each user and categorize each into groups of 5-10 tabs, more than 15 tabs, 10-15 tabs. I converted to excel to make a pie chart showing the distribution of max number of tab groups. While the majority have never gone above 5 tabs, the most people have had 5 to 10 tabs with a group of outliers having more than 10 tabs going. out of 14,000 users, only about 700 of them have used more than 10 tabs at once.



- I wanted to test my theory that people who access firefox via mobile tend to use a higher number of tabs than people who don't access via mobile.
    a. I had to do 3 queries to find this: select avg(rtrim) from(select distinct user_id, cast(trim(trailing 'tabs' from data2) as int), q9 from events inner join survey using(user_id) where event_code = 26) as temp where q9 like '%3%'
        i. this told me that the average firefox user who used via mobile had 12.9 tabs open
    b. Second query: select avg(rtrim) from(select distinct user_id, cast(trim(trailing 'tabs' from data2) as int), q9 from events inner join survey using(user_id) where event_code = 26) as temp where q9 not like '%3%'
        i. This said that people who did not access via mobile averaged 8.8 tabs total
    c. Then i just checked the average amount of tabs the average user was using, (basically took out the final where statement)
        i. This said that the average user used 10.3 tabs. So this means that the average user who used mobile used 20% more tabs than people who did not use mobile.

Do you agree or disagree with your opinionated colleague?

Looking at the data, you would see that the Firefox users represented use a much larger volume of bookmarks on average than tabs. Rarely, would they ever go above 10 tabs at max. So I would disagree with my opinionated colleague. If I were to make edits to the support of tabs, I would focus on the mobile platform, which the users who access via mobile tend to use more tabs than people access via computer.

Do you think the team should go in a different direction?

Instead of thinking of the question as bookmarks or tabs, I think it would make more sense to think about educating firefox users on the different capabilities and ease of use with the bookmarks to the newer users, while focusing on optimizing tabs for mobile users.

Include analysis of data with statistics and outliers.
In master excel grid

Present any limitations and assumptions.
Overall, the data tends to skew more towards audiences with a high skill level of internet/computer usage, so this could present a limitation of the data since it's not quite representative of all the users. Also, It would be nicer if the engineering department helped to do the following to support future endeavors:

- Clarify the device that the firefox session was used on.
- Maybe information about the age of bookmarks in the folder and the amount of times that bookmark is referred to

Key Assumptions:

- It's better for business to establish a good experience for new users than it is for older users.