Which parts of an article are actually being read?

Tilman Bayer

Wikimania 2018, Cape Town, July 22, 2018

WIKIMANIA
CAPE TOWN

Beyond pageviews

What knowledge do people access on Wikimedia projects, and how often?
Pageview numbers have been our main source for answering this question.
However, they only tell us how often a page is opened - not how much of it is being read.

This talk

WIKIMANIA
CAPE TOWN

- Tour through some new data sources that became available in recent years
- These come from academic research or from data WMF collected during work on software features (products). But they yield insights for editors too.
- Myself: Analyst in the WMF Audiences (product) department, Wikipedia editor since 2003

Data on scroll actions

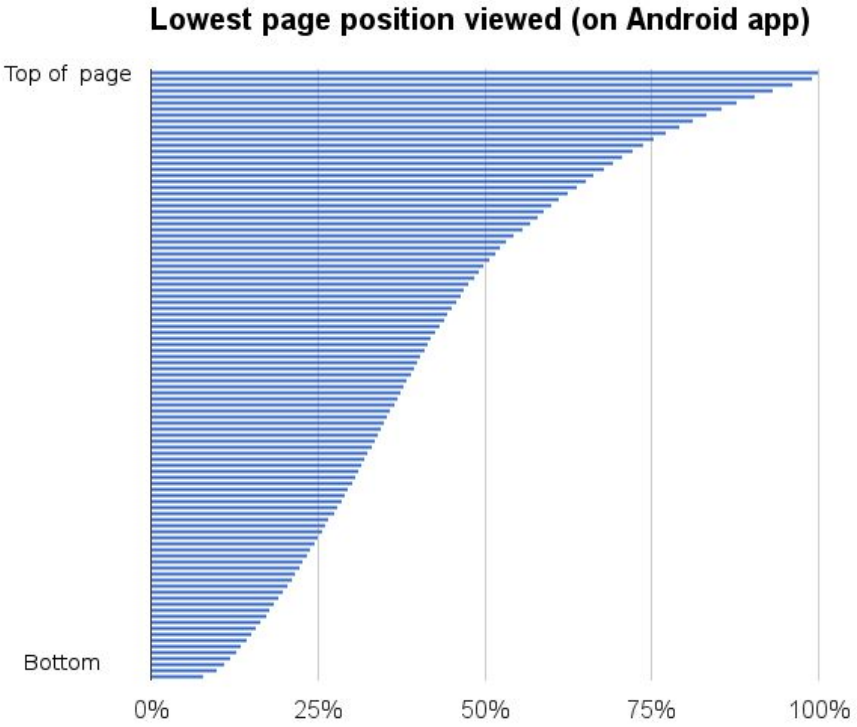
On the Android Wikipedia app,

Around 32% of pageviews do not involve scrolling down - i.e. reader looks only at the top of the page.

Scrolling data

On the Android Wikipedia app,

- **The bottom of the page is only reached on about 5% of pageviews**
- **Only 25% of users ever reached the end of any page (during one week)**

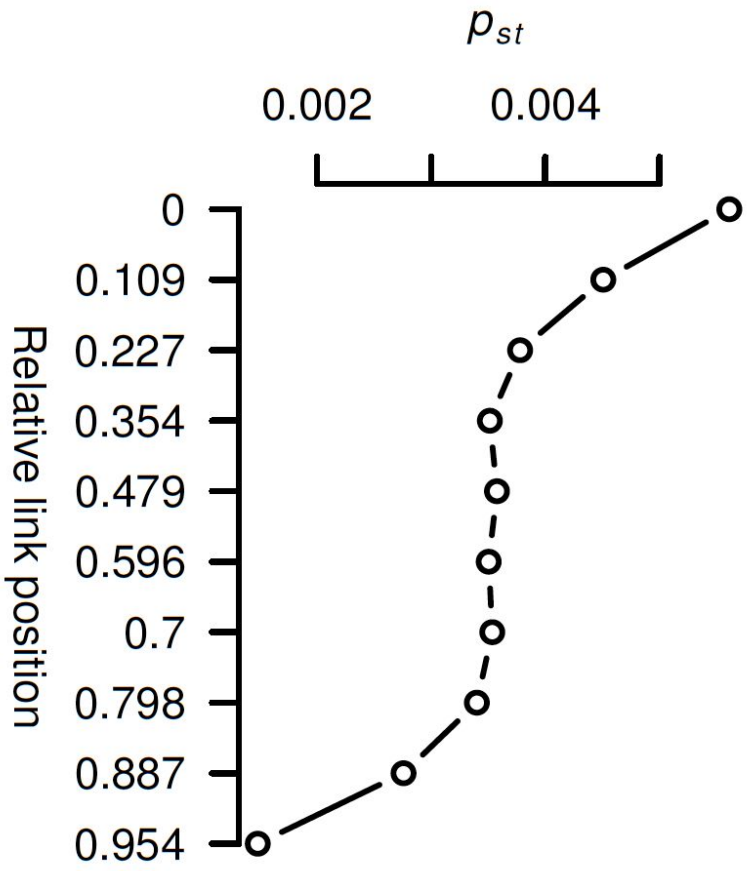


Clickstream data

Public dataset by WMF, counting how often wikilinks are followed from one page to the other.

Likewise indicates that the top of an article is read much more often that its end.

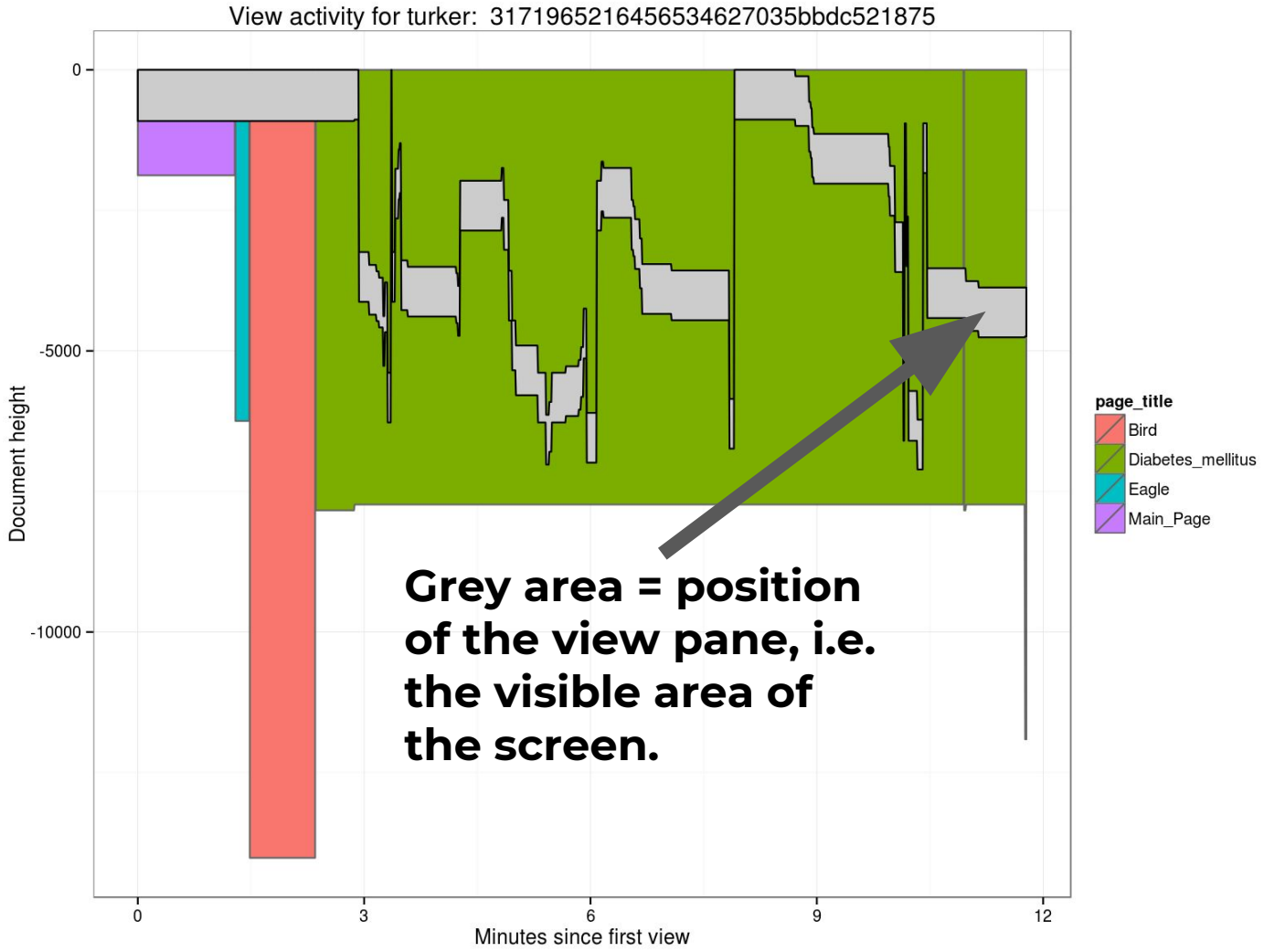
WIKIMANIA CAPE TOWN



Clickthrough rate of wikilinks vs. their position on the source page, 0=top, 1=bottom (from Paranjabe et al.)

Pages may not be read linearly.

Here, the reader moved through several Wikipedia articles, jumping up and down, while looking for a specific fact about diabetes.
[(Mechanical Turk experiment by Halfaker et al., 2013)]



Eyetracking studies

Expensive lab equipment allows to find out which part of the screen a user looks at.

WMF has never used this so far, but several academic researchers (from Germany and Scotland) have applied it to (desktop) Wikipedia.

Eyetracking

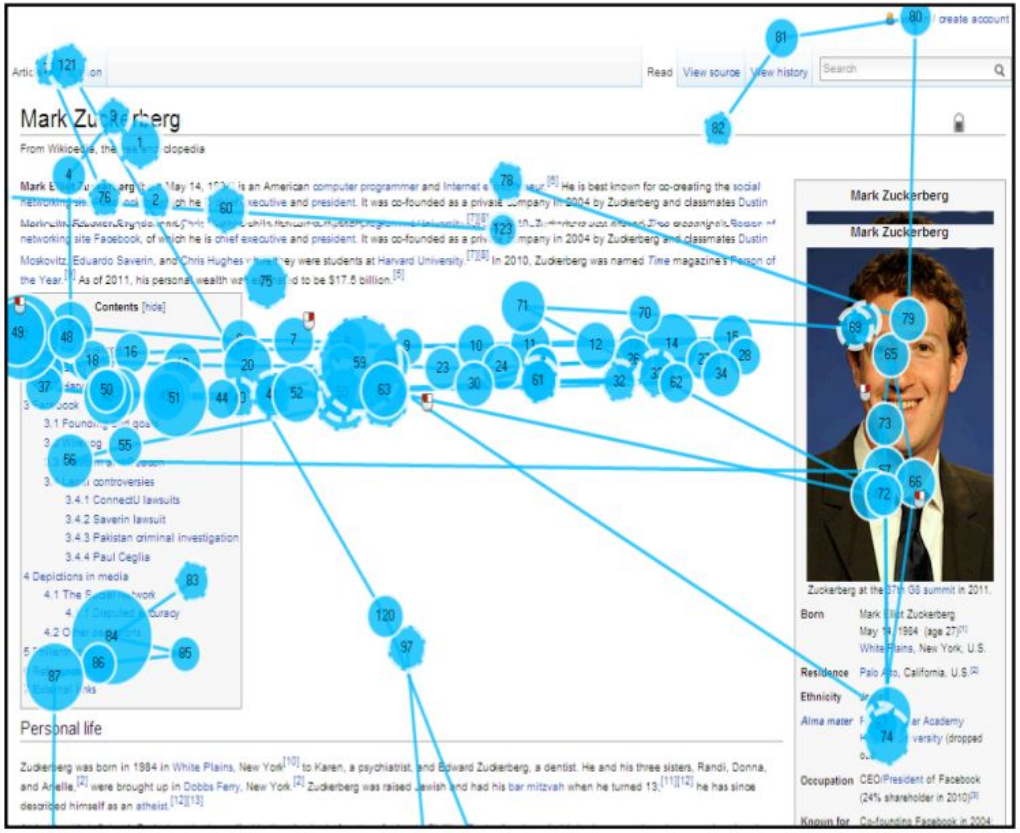


Figure 1. Gaze plots example from Mark Zuckerberg article. The dots are fixations, larger dots mean greater fixation durations and the lines between the fixations are saccades.

Insights from eyetracking studies

- Most readers look first at TOC, then infobox.
- In "lookup" tasks, >45% of time is spent scanning TOC and lists.
- In "learn" tasks, it's <10%.
- Images provide visual orientation, too:
  - “39% of all [eye] contact points lead from image to image, in mutual directions (previous or next).“

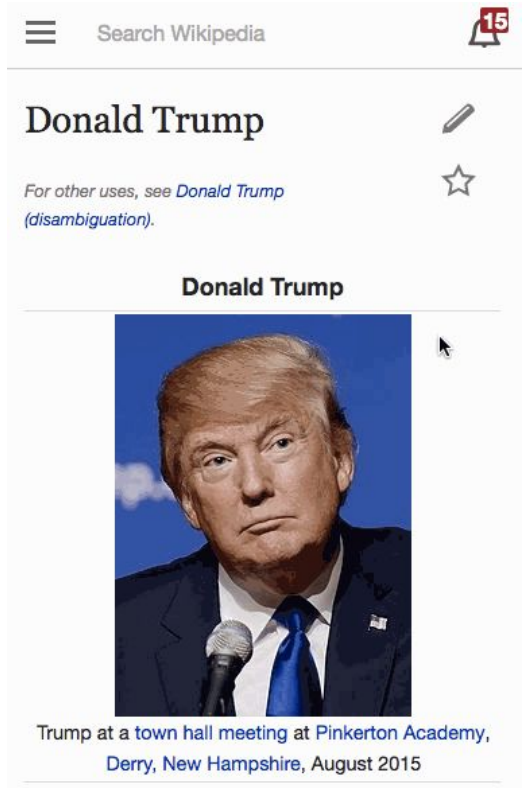
(sources)

Section expansions

On the mobile web version of Wikipedia, article sections are collapsed until the reader taps the heading

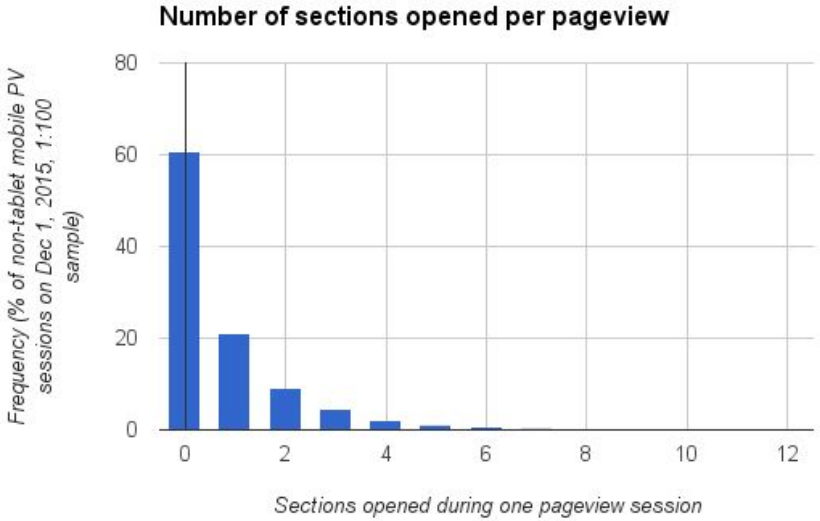
We instrumented this for an A/B test, and now have some data about how often each section is accessed

WIKIMANIA CAPE TOWN



(click for animated screencast)

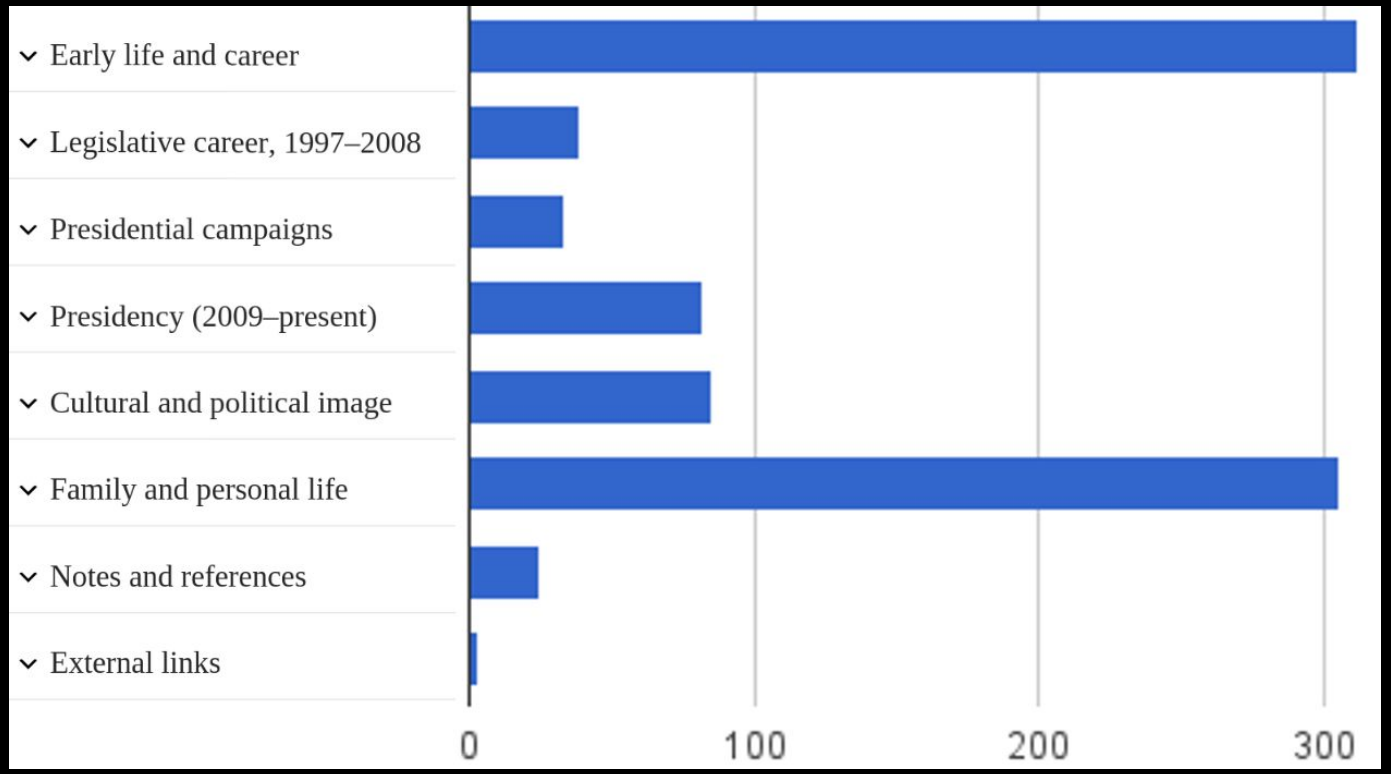
On 60% of mobile pageviews, not a single section is opened, i.e. the reader only looks at the introduction.



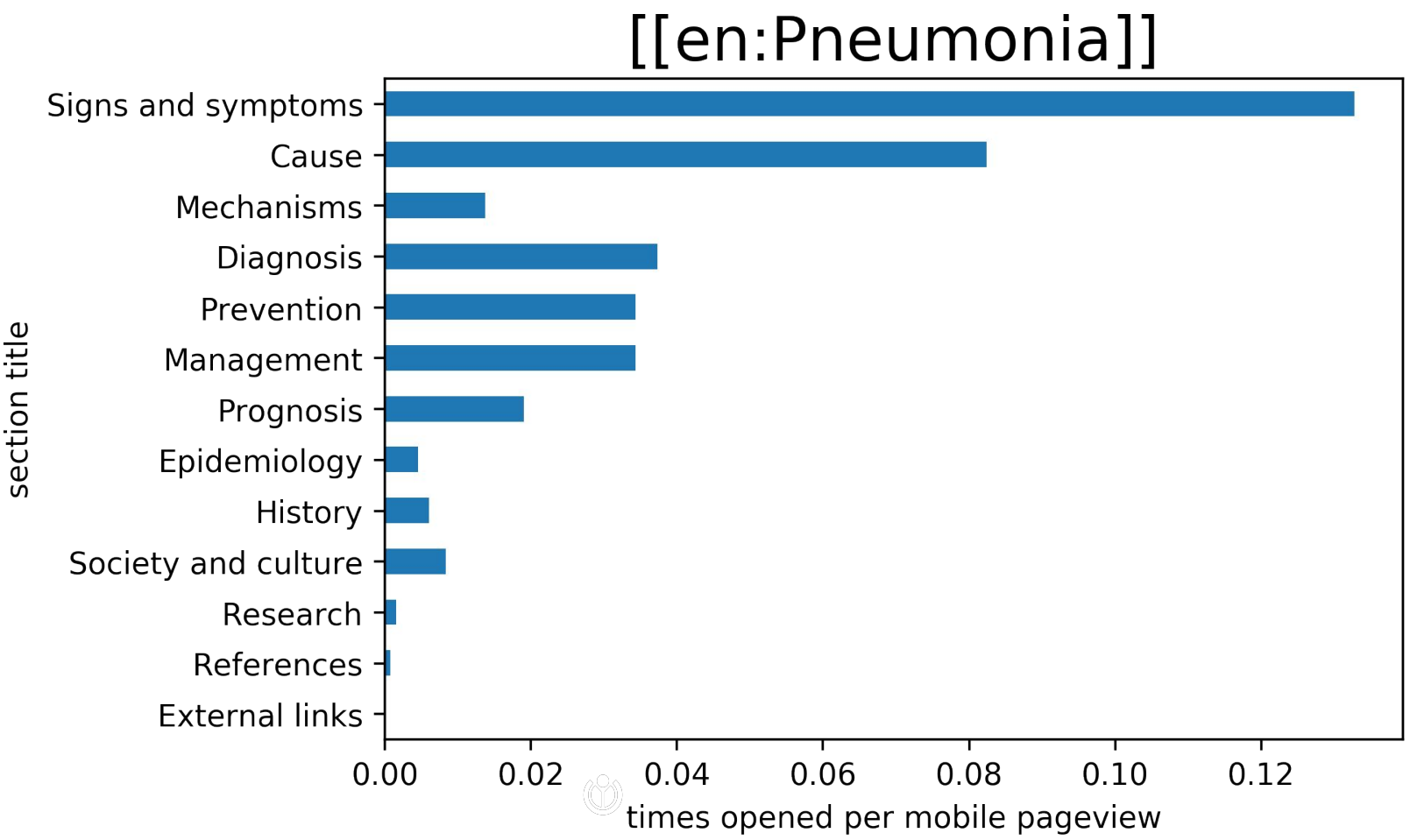
Much more interest for personal aspects (“Early life and career”) than for political work (“Legislative career”, “presidency”)

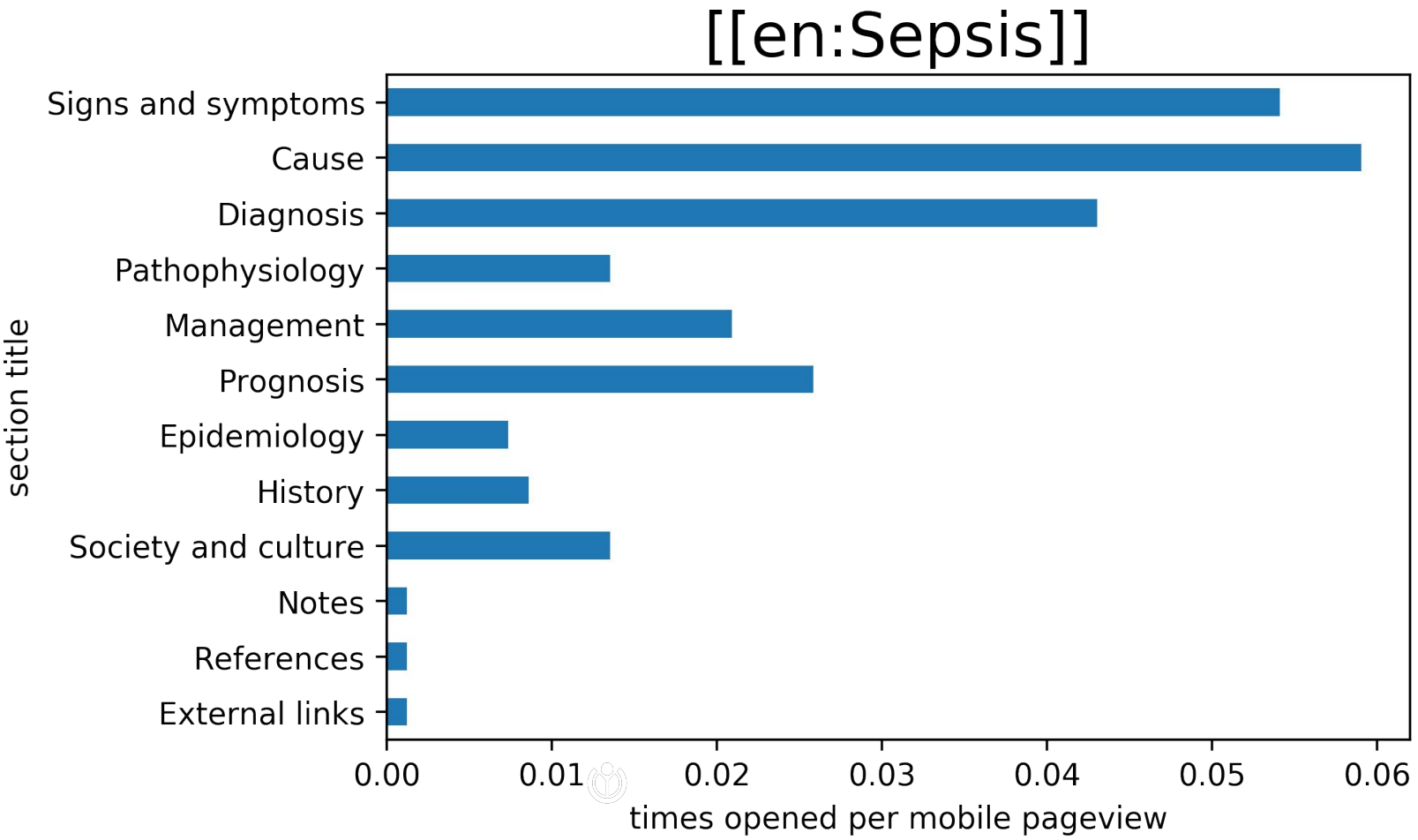
Almost no-one views “external links”

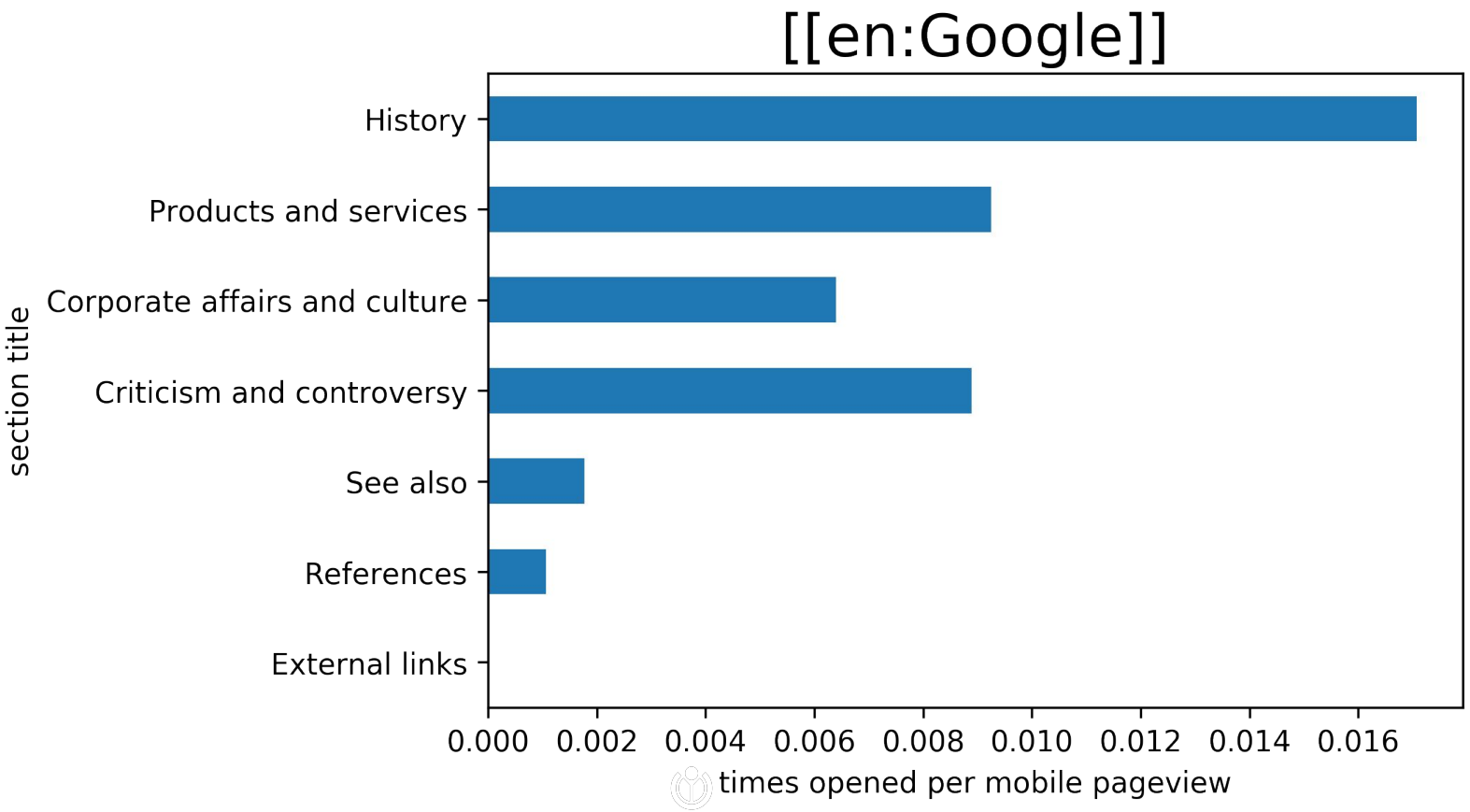
Article “Barack Obama” (English Wikipedia)



Times opened in sample (Dec 2015-Jan 2016)


[[en:Pneumonia]]
Signs and symptoms
Cause
Mechanisms
Diagnosis
Prevention
Management
Prognosis
Epidemiology
History
Society and culture
Research
References
External links
section title
0.00
0.02
0.04
0.06
0.08
0.10
0.12
times opened per mobile pageview


[[en:Sepsis]]
Signs and symptoms
Cause
Diagnosis
Pathophysiology
Management
Prognosis
Epidemiology
History
Society and culture
Notes
References
External links
section title
0.00
0.01
0.02
0.03
0.04
0.05
0.06
times opened per mobile pageview


[[en:Google]]
History
Products and services
Corporate affairs and culture
Criticism and controversy
See also
References
External links
section title
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
0.016
times opened per mobile pageview

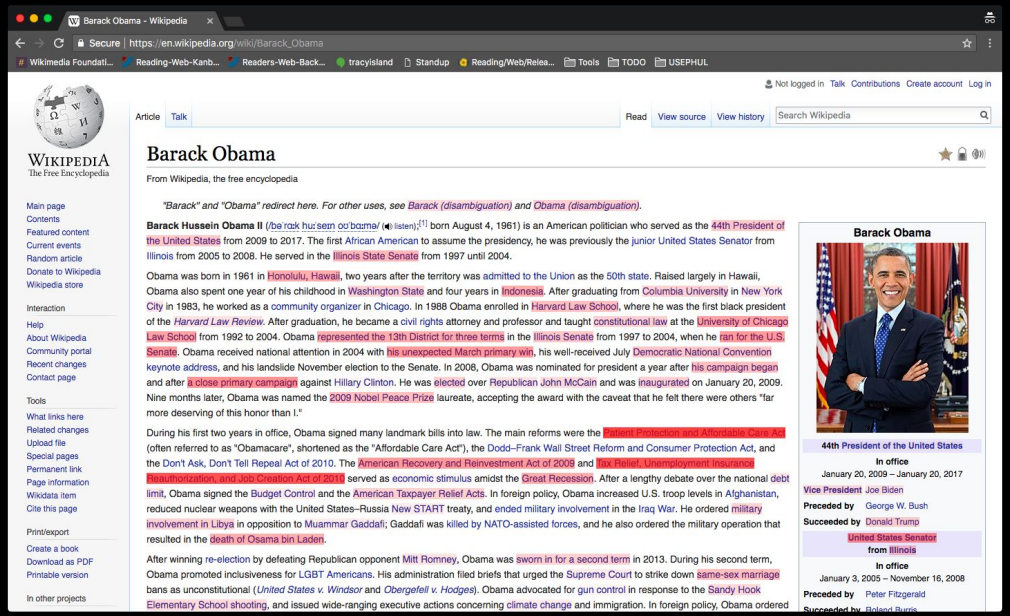
Some takeways from these and other pages

- Mobile web readers seem very uninterested in the “external links” section.
- “References” is rarely opened as well (but footnotes can also accessed from the text).
- Lower section less popular, but heading/topic seems to matter.

Page previews data

- New feature (2017/18) on desktop Wikipedia
- Reader hovers over a link to see an excerpt of the linked data
- Heatmap of hovers shows which parts of the page were read, but also which topics(links) readers are interested in looking up briefly from the source page

Page previews data



WIKIMANIA CAPE TOWN

Thank you!
Questions?

Tbayer at wikimedia dot org

[[meta:Research:Which parts of an article do readers read]]

WIKIMANIA
CAPE TOWN