

R Rising? Exploring Technology Trends in Data Journalism

Erica Yee | yee.er@northeastern.edu

Introduction

- Newsrooms are increasingly incorporating the gathering, analyzing and presenting of data into their reporting (Coddington 2015).
- This study discovers technology trends in the evolving **data journalism** field.

Research Questions

- What **technologies and tools** are used in the practice of data journalism? How have they changed over time?

Short Answer

- Findings suggest **Microsoft Excel, Python, SQL, R and JavaScript** have been used consistently by data journalists over the last five years, while other technologies have fallen out of favor.
- R especially has seen a recent **surge in popularity**, which could have implications for practitioners from students to newsrooms.

Approach

- The data came from the National Institute for Computer Assisted Reporting (NICAR), under the assumption that topics discussed at the major conference in the field reflect what practitioners use and where the industry is going. NICAR is an important organization for data journalism practitioners, providing resources and connecting professionals and students (Fink & Anderson 2015).
- Five years (2015-2019) of **conference session descriptions** were analyzed.
- Types of **keywords extracted** to assess popularity over time included programming languages, libraries, frameworks and software.
- The collected and categorized data enables **time trend analyses and recommendations** based on the current technology landscape.

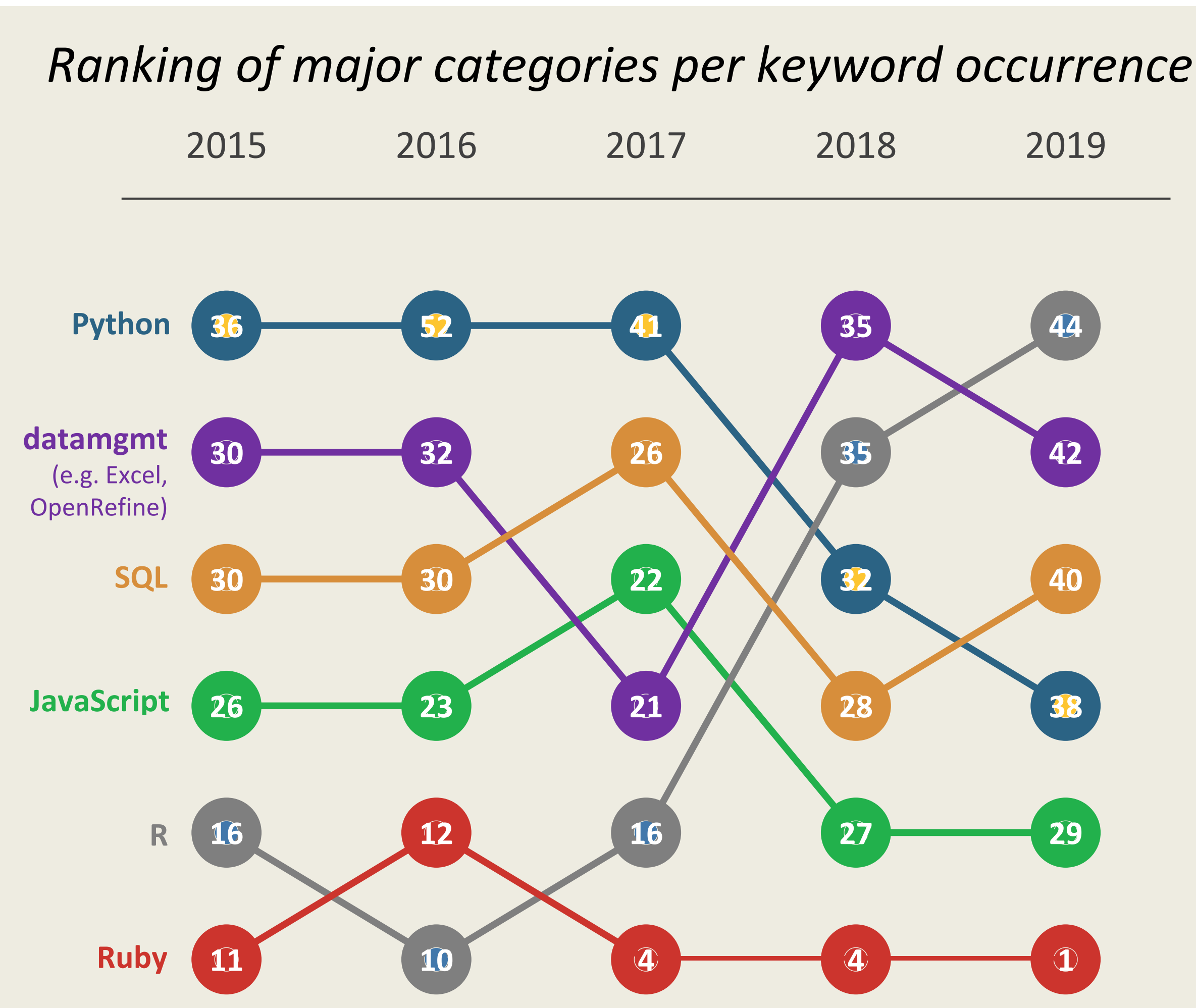
Impact

- The gleaned technology landscape offers **practical suggestions** for those wanting to explore technologies for data journalism.
- This table of top technologies from 2019 can be a useful starting place for both students and professionals.

Category	Top technologies mentioned in 2019
<i>Python</i>	Jupyter Notebooks, matplotlib, pandas
<i>JavaScript</i>	D3.js
<i>R</i>	ggplot2, RStudio, tidyverse
<i>web</i>	HTML, CSS
<i>SQL</i>	MySQL, SQLite
<i>datamgmt</i>	Excel, pivot tables
<i>geo</i>	QGIS
<i>dataviz</i>	Tableau
<i>stats</i>	PSPP, SPSS
<i>other</i>	GitHub

Results

In total, 117 keywords were extracted and labeled into 15 categories.

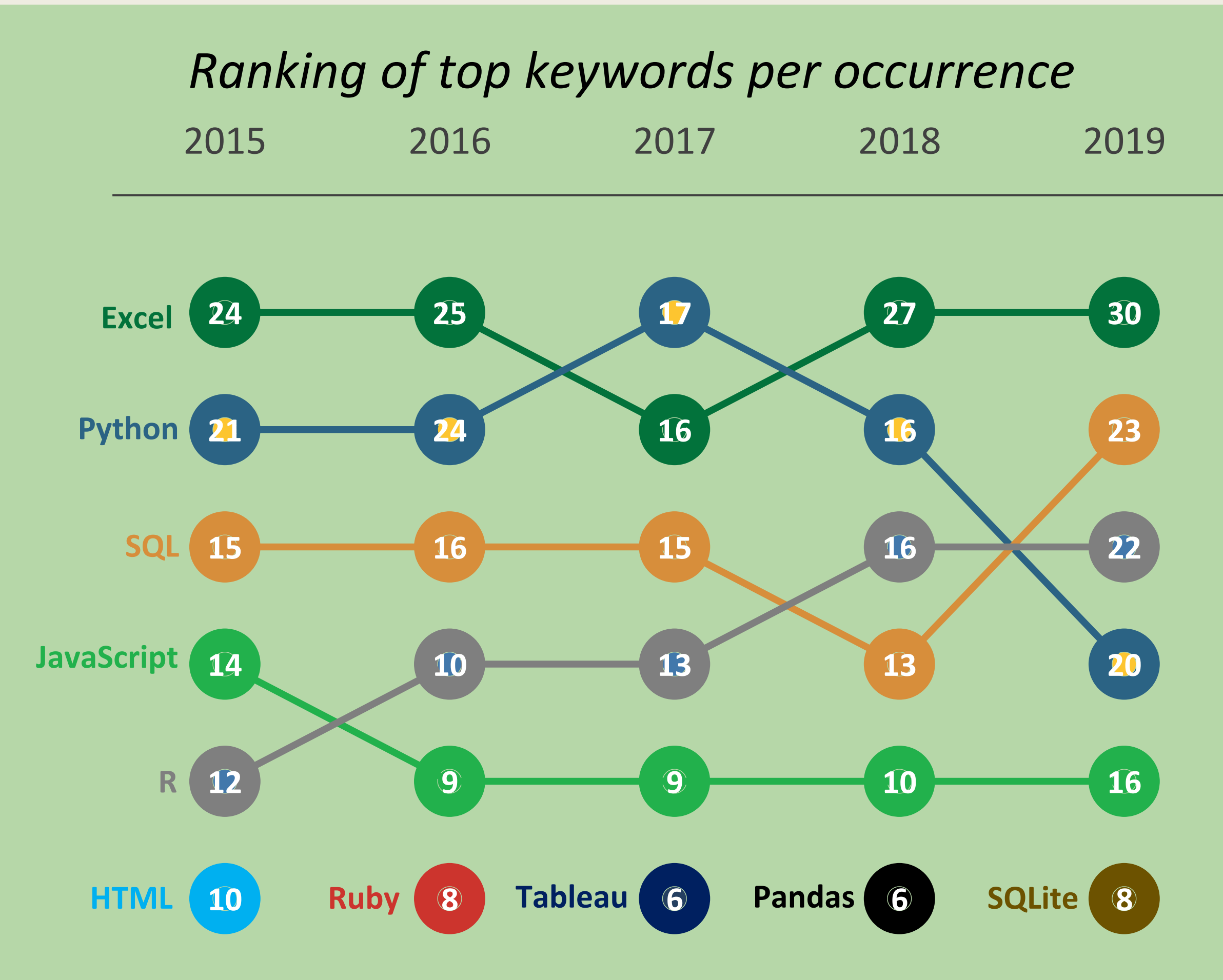


The Python category had the most keywords overall of any programming language. SQL and JavaScript keywords occurred consistently, while the **number of R-related keywords jumped dramatically in 2018**. The “datamgmt” category, which includes data management and wrangling tools like Microsoft Excel, ranked consistently high.

The top 5 keywords remained in that range from 2015-2019, while the sixth-ranked changed every year.

Discussion

- The breadth of technologies suggest a **wide variety** used in practice for data and digital journalism. The diversity of tools indicate many paths to similar end goals.
- While certain tools fall out of favor and use, some seem to remain the **prevailing standard like Excel**, which is relatively widespread and accessible.
- Amid the perpetual struggle between Python and R for data analysis, **R's apparent rise** may confirm its usefulness in this field. Still, further research and time is needed to corroborate this finding.



Key References

Coddington, M. (2015). Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 331-348.

Fink, K., & Anderson, W. (2015). Data Journalism in the United States. *Journalism Studies*, 467-481.

Acknowledgements

Professor Martin Schedlbauer

Professor John Wihbey

Professor Matt Carroll

Laura South