

# Discovery of Trends with Tags of Stack Overflow

...

December 18, 2018

# Overview

- Stack Overflow has many groups separated by interests
- Search of Stack Overflow is facilitated by proper topic tagging
- Stack Exchange can be used to query Stack Overflow tags (and explore their popularity)
- Russian Stack Overflow is also a large and active community
- Comparing tags of the English-speaking Stack Overflow to the Russian counterpart may reveal interesting trends

# Motivation

- As an IT professional, I use Stack Overflow frequently
- Sometimes there are answers on the Russian site, which cannot be found within the English-speaking community
- Data Camp has a project, which explores Stack Overflow tags as a way of finding the most popular programming languages
- However, the Data Camp project lacks:
  - R Shiny app to facilitate interactive trend discovery and
  - tag analysis of the Russian-speaking community

# Null and Alternative Hypotheses:

## Null Hypothesis

Russian and English tags of Stack Overflow belong to the same population group (Russian Stack Overflow is simply a smaller subset of the same tag space).

## Alternative Hypothesis

Russian and English tags of Stack Overflow represent two different population groups (with distinct tags and popularity).



## Choose a Site

Data updated yesterday



### Stack Overflow

Q&A for programmers

17m  
questions

26m  
answers

70m  
comments

54k  
tags

[visit site →](#)



### Mathematics

Q&A for people studying math at any level and professionals in related fields

1m  
questions

1.4m  
answers

4.5m  
comments

1.8k  
tags

[visit site →](#)



### Super User

Q&A for computer enthusiasts and power users

392k  
questions

580k  
answers

1.4m  
comments

5.3k  
tags

[visit site →](#)



### Ask Ubuntu

Q&A for Ubuntu users and developers

312k  
questions

397k  
answers

1.1m  
comments

3k  
tags

[visit site →](#)



### Server Fault

Q&A for system administrators and IT professionals

269k  
questions

442k  
answers

886k  
comments

3.6k  
tags

[visit site →](#)



### Stack Overflow на русском

Q&A for программистов

245k  
questions

282k  
answers

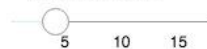
1m  
comments

4.4k  
tags

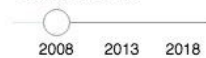
[visit site →](#)



Number of Tags:

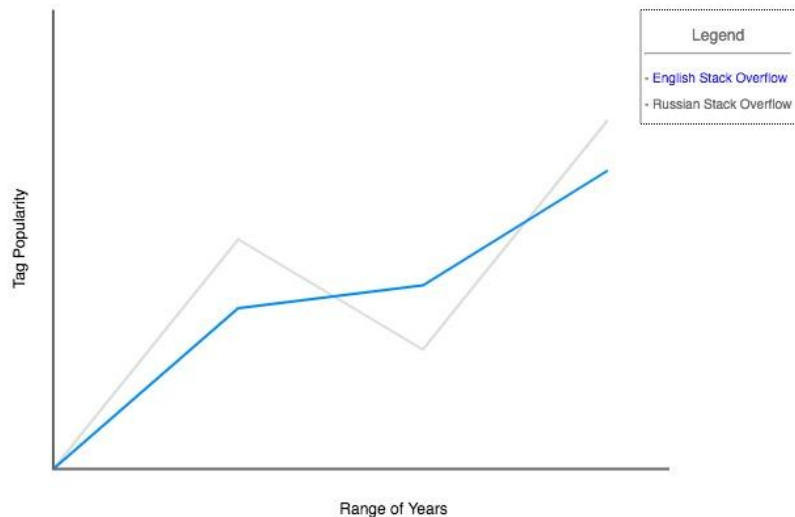


Range of Years:



Autocomplete Tags:

Popularity of Tags in Russian vs. English Stack Overflow

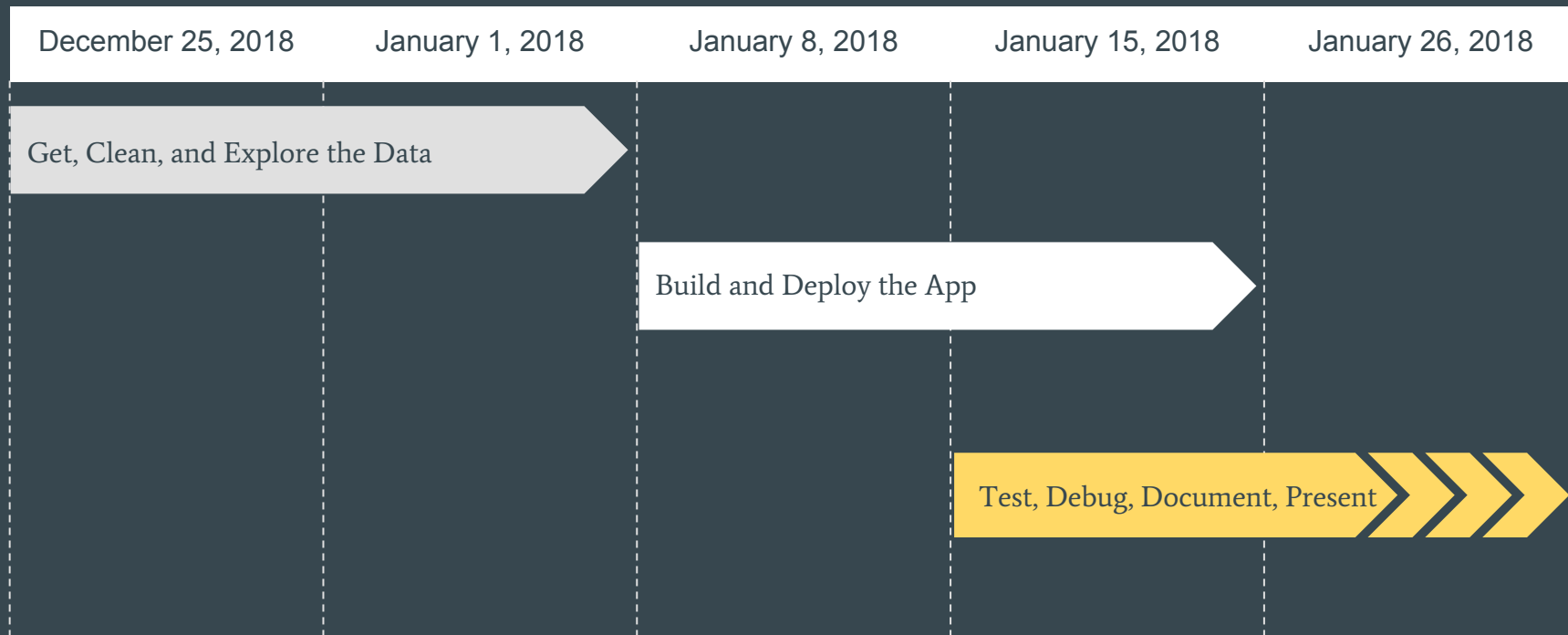


# Evaluation

Initial deliverable will facilitate ***visual*** inspection and discovery of trends.

---

# Timeline





# Issues and Challenges

- Querying Stack Exchange and tidying data for analysis
- Getting enough data to compare trends
- Addressing possibility of missing data and difficulty of cross-language tag comparison
- Testing user interface with volunteers
- Debugging the application to address revealed shortcomings
- Evaluating insight usefulness and suggesting future improvements

