

Prezentacja wyników pierwszego projektu

Komosiński D, Krubiński M, Kurek M

18 kwietnia 2018



Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska



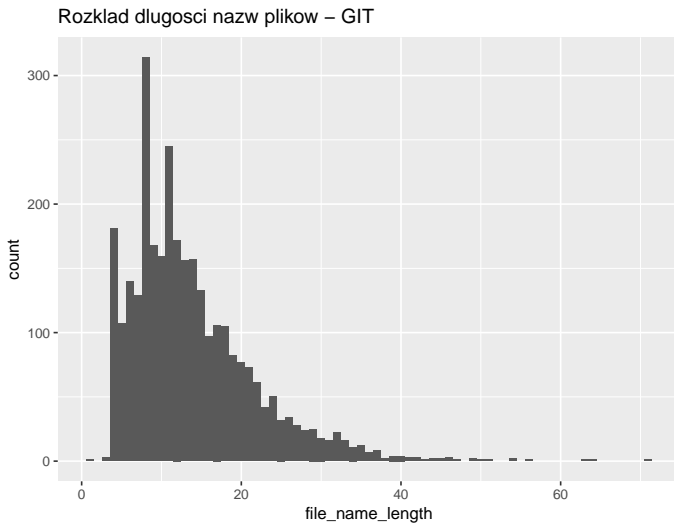
200 użytkowników z GIT, zwróconych przez API dla zapytania o użytkowników, którzy mają pomiędzy 50 a 100 repozytoriów, w tym repozytoria w R. Wyszło ok. 3000 skryptów

Cały CRAN, wydobywaliśmy pliki DESCRIPTION i zawartość folderu R. Przeanalizowaliśmy wyniki całościowe i częściowe. W częściowych wzięliśmy pod uwagę 1700 bibliotek, a znajdowało się w nich 23 000.

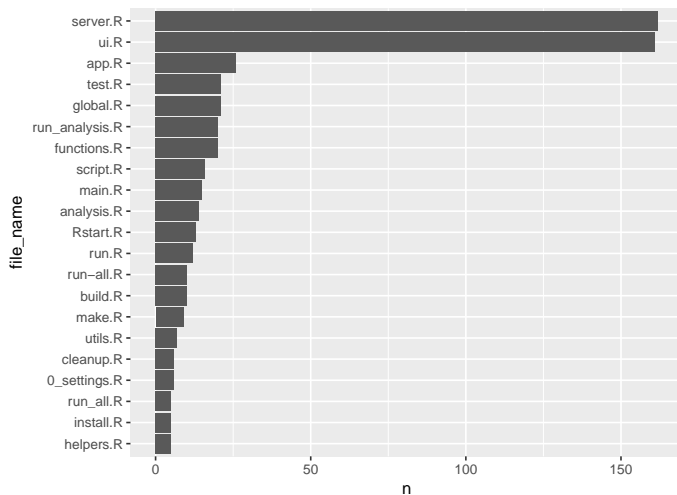
Badaliśmy:

- długość nazw zmiennych i funkcji
- długość nazw plików
- licencje
- konwencję nazywania zmiennych i funkcji
- zaimportowane pakiety

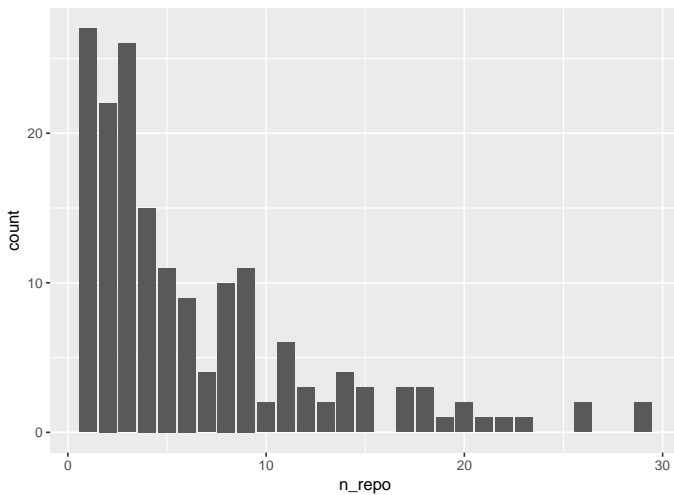
GitHuB



Najczęstsze nazwy plików – GIT

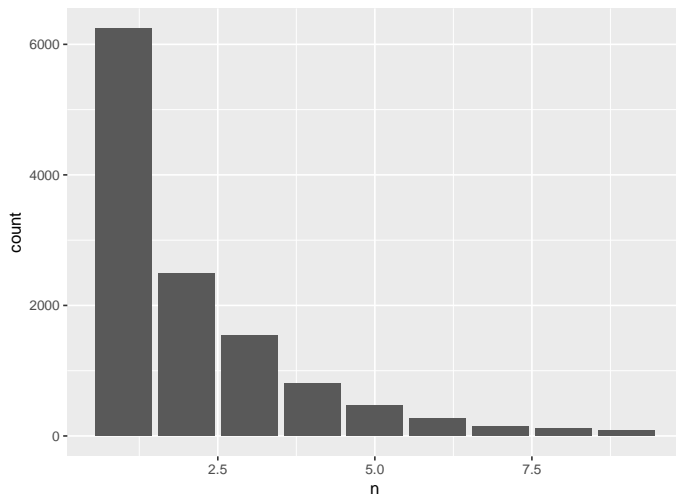


Histogram licznosci repozytoriow per autor – GIT

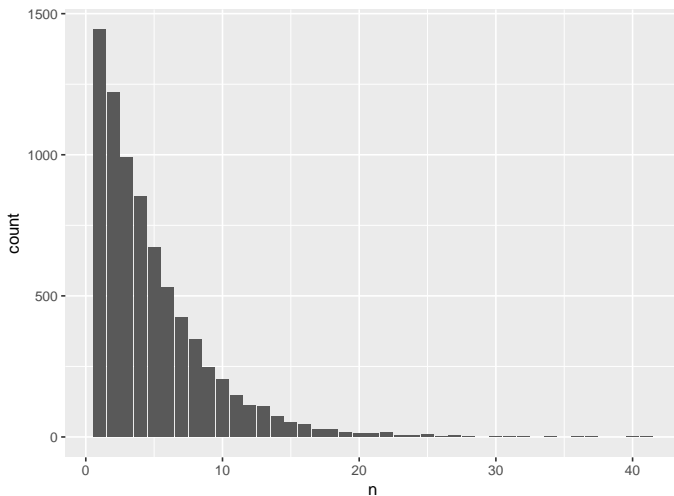


CRAN

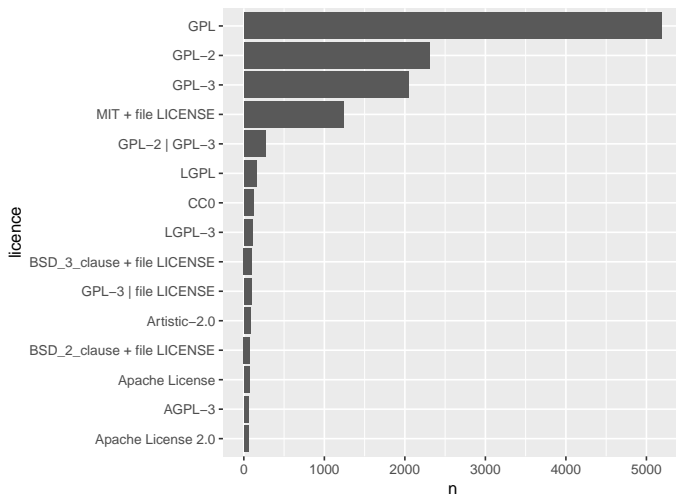
Histogram liczby autorów per pakiet – CRAN



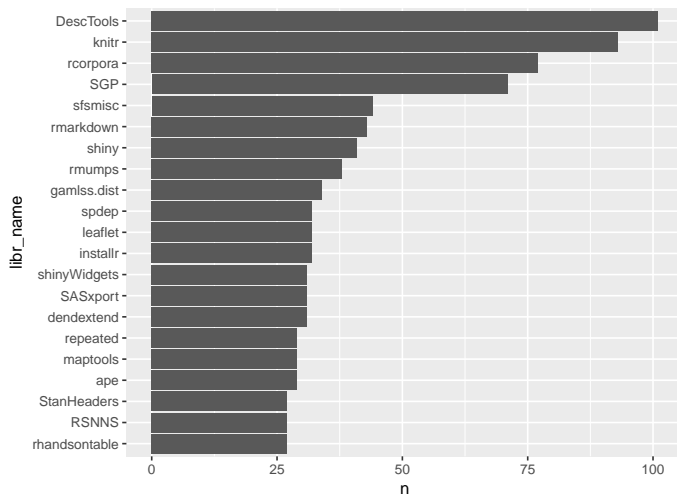
Histogram liczności użytych pakietów per biblioteka – CRAN



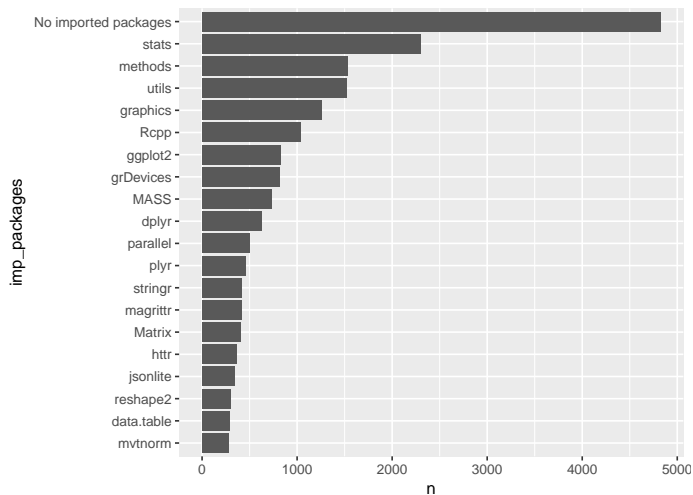
Najczęściej używane licencje – CRAN



Pakiety o największej liczbie autorów – CRAN

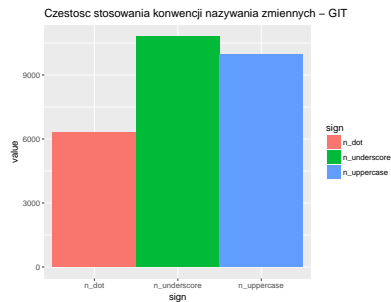
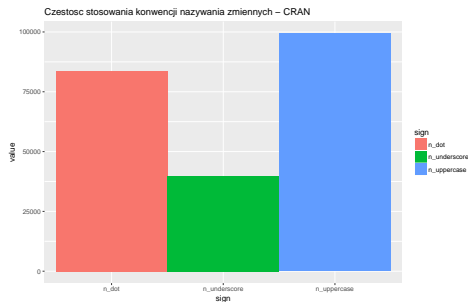


Najczęściej stosowane pakiety – CRAN



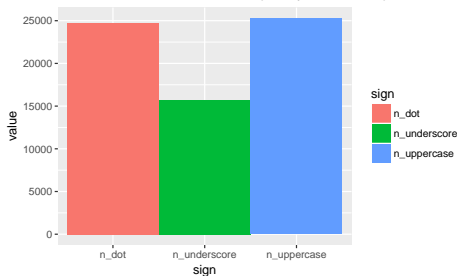
Porównania GitHuba i CRANa

Częstość stosowania konwencji nazywania zmiennych

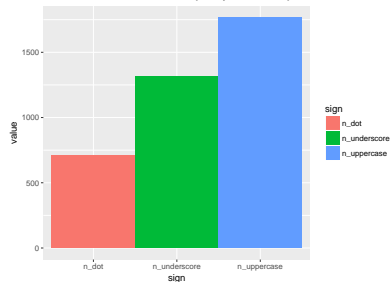


Częstość stosowania konwencji nazywania funkcji

Częstość stosowania konwencji nazywania funkcji – CRAN

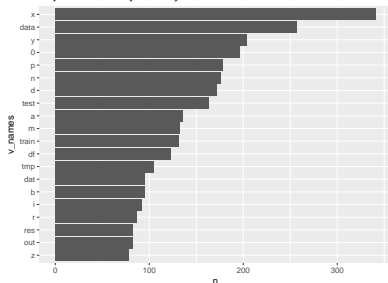


Częstość stosowania konwencji nazywania funkcji – GIT

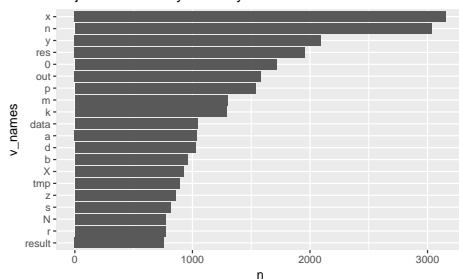


Najczęstsze nazwy zmiennych

Najczęstsze nazwy zmiennych – GIT

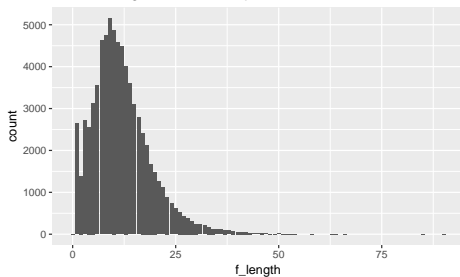


Najczęstsze nazwy zmiennych – CRAN

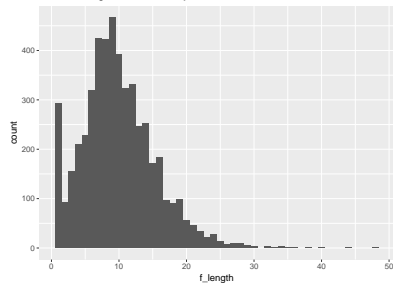


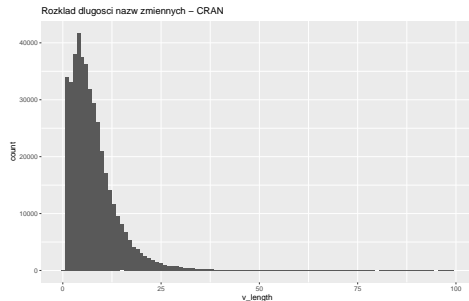
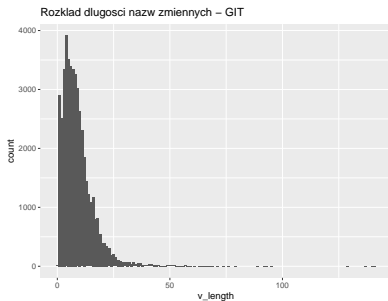
Najczęstsze nazwy zmiennych

Rozkład długości nazw funkcji – CRAN



Rozkład długości nazw funkcji – GIT





```
relative_probs_of_each_state_at_branch_top_AT_node_DOWNPASS
```

```
calc_loglike_sp_stratified_results$ML_marginal_prob_each  
_state_at_branch_bottom_below_node
```

Dziękujemy za uwagę!