

---

# **Analiza danych historycznych o bitwach i zawartych pokojach w oparciu o angielską wikipedię**

---

Piotr Szmigielski, Marcin Lis

---

# Wstęp

---

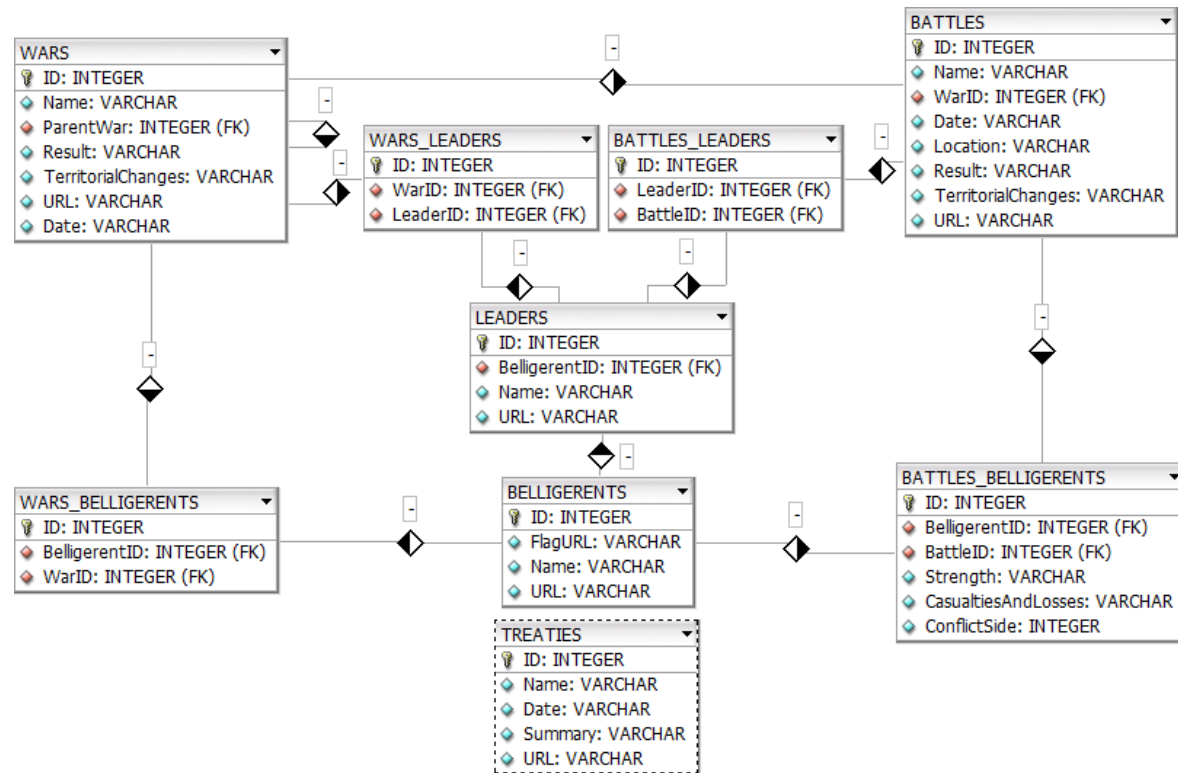
- Cel: pozyskanie i analiza danych o bitwach, wojnach i pokojach
  - Zakres czasowy: lata 1301-1800
  - Podział na 3 etapy: pobranie danych, przetworzenie danych i wizualizację rezultatów
  - Źródło danych: angielska wikipedia
-

# Technologie

---

- Język programowania: C#
  - Baza danych: relacyjna baza MS SQL
  - Pozyskiwanie danych i parsowanie:  
biblioteka HTML Agility Pack
  - Persystencja: NHibernate
  - Wizualizacja: Gephi, LibreOffice Calc  
(wykresy)
-

# Schemat bazy danych



# Etapy działania

---

- Pobieranie danych
    - bitwy, wojny z [en.wikipedia.org/wiki/List\\_of\\_battles\\_1301-1800](https://en.wikipedia.org/wiki/List_of_battles_1301-1800)
    - pokoje
  - Przetwarzanie danych
  - Wizualizacja i analiza danych
-

# Przetwarzanie danych

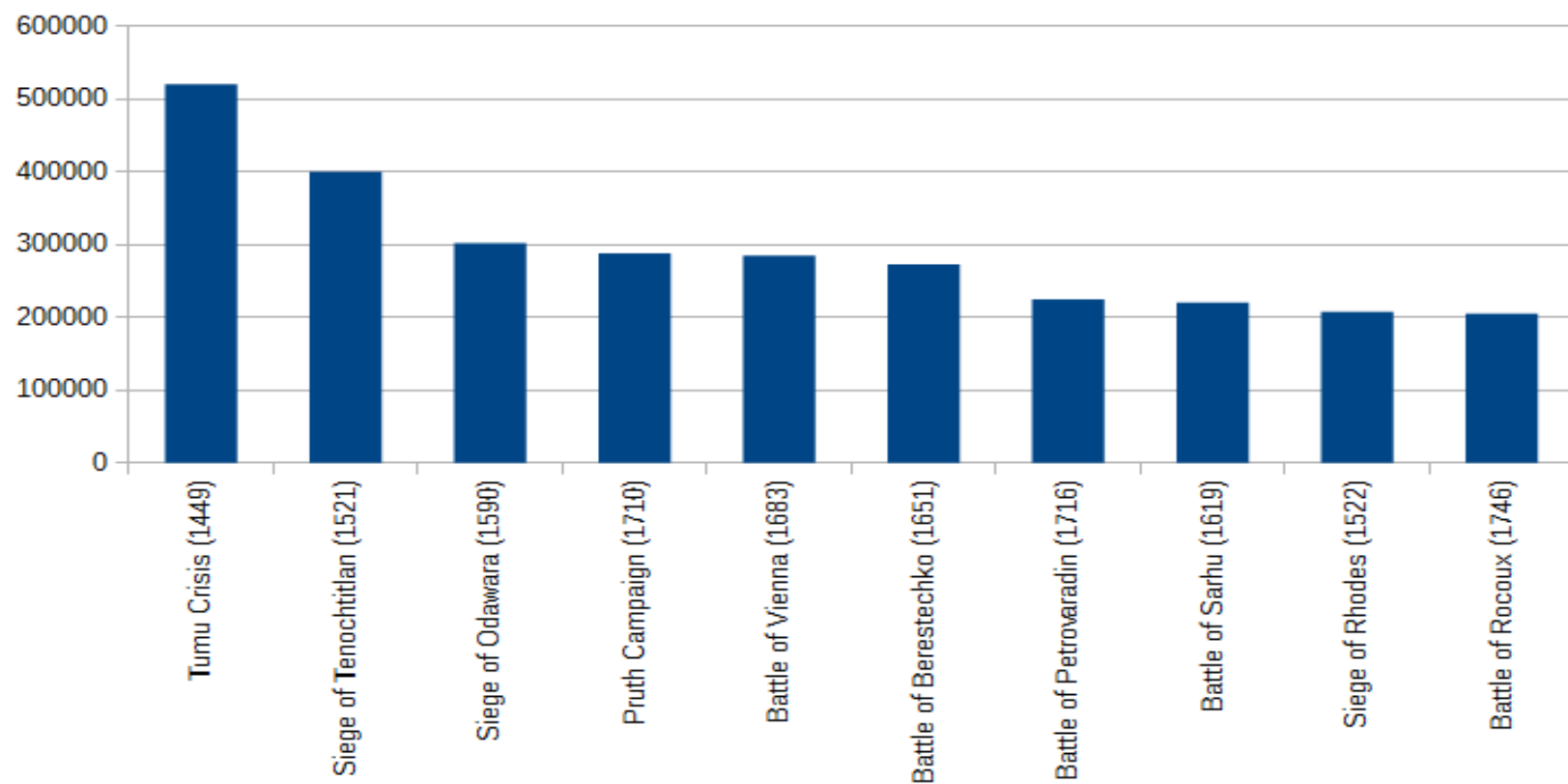
---

Następujące dane pozyskiwane z pól tekstowych:

- Data rozpoczęcia i zakończenia bitwy 70%
  - Rodzaje sił w bitwie (piechota, kawaleria, artyleria, flota, suma) 60%
  - Straty w bitwie (zabici, ranni, w niewoli, suma, zatopione statki) 70%
  - Dzisiejszy kraj na terenie którego toczyła się bitwa 55%
  - Rezultat bitwy/wojny (Decisive Win, Win, Lose, Decisive Lose, Inconclusive) 60%
  - Grupa strony konfliktu 0%
-

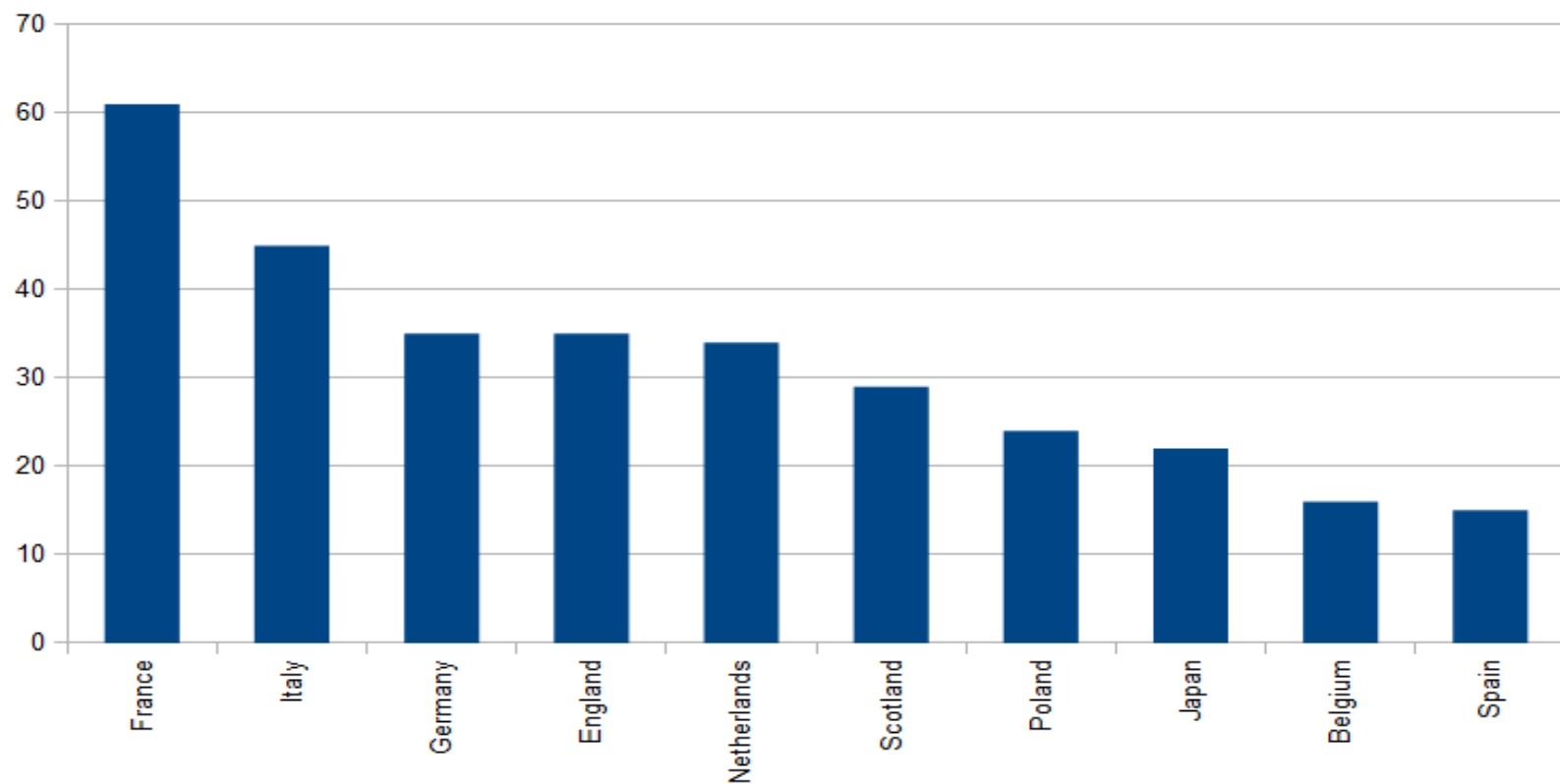
## Suma sił uczestniczących w bitwie

top 10



## Liczba bitew na terenach dzisiejszych państw

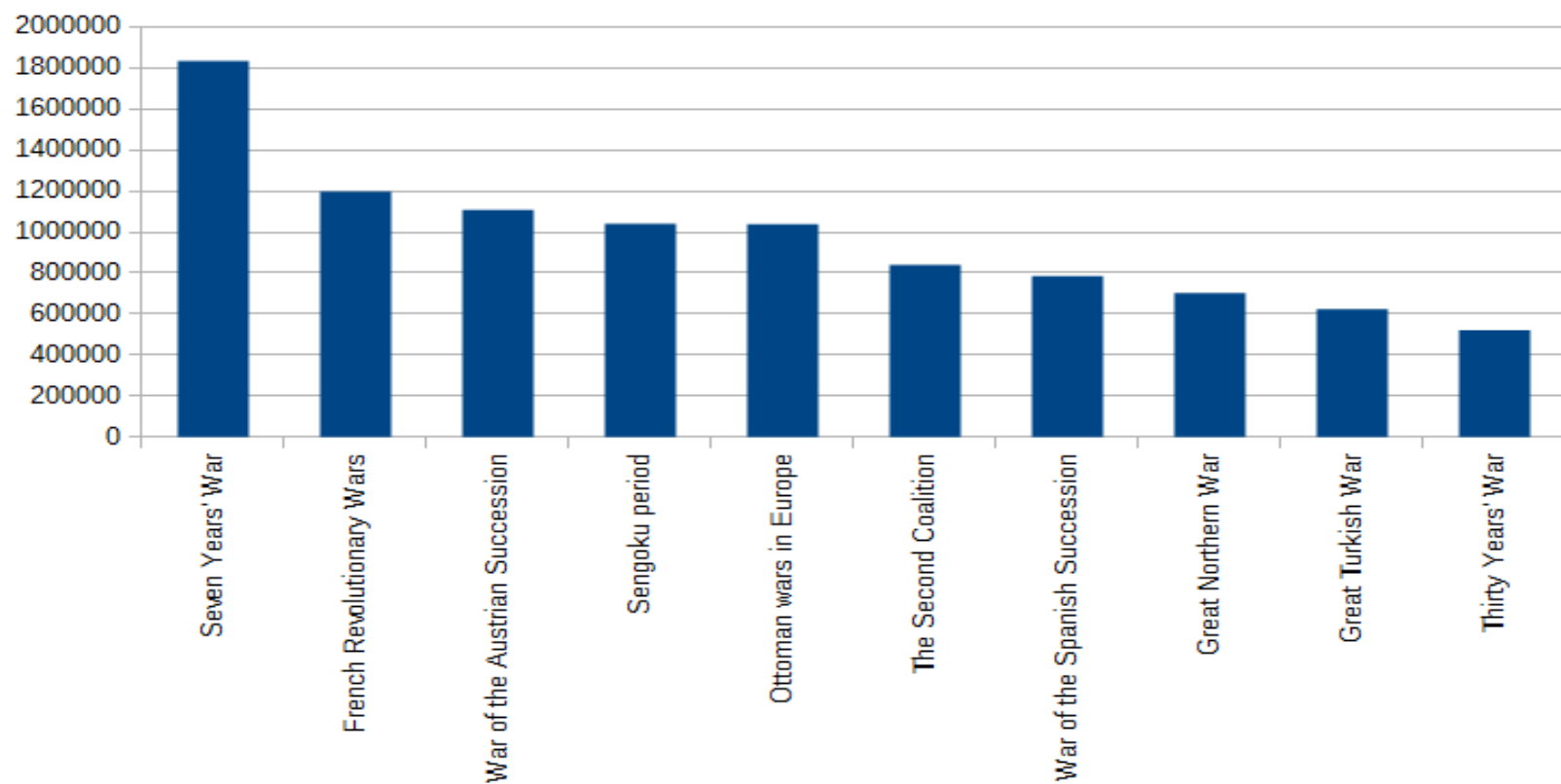
top 10





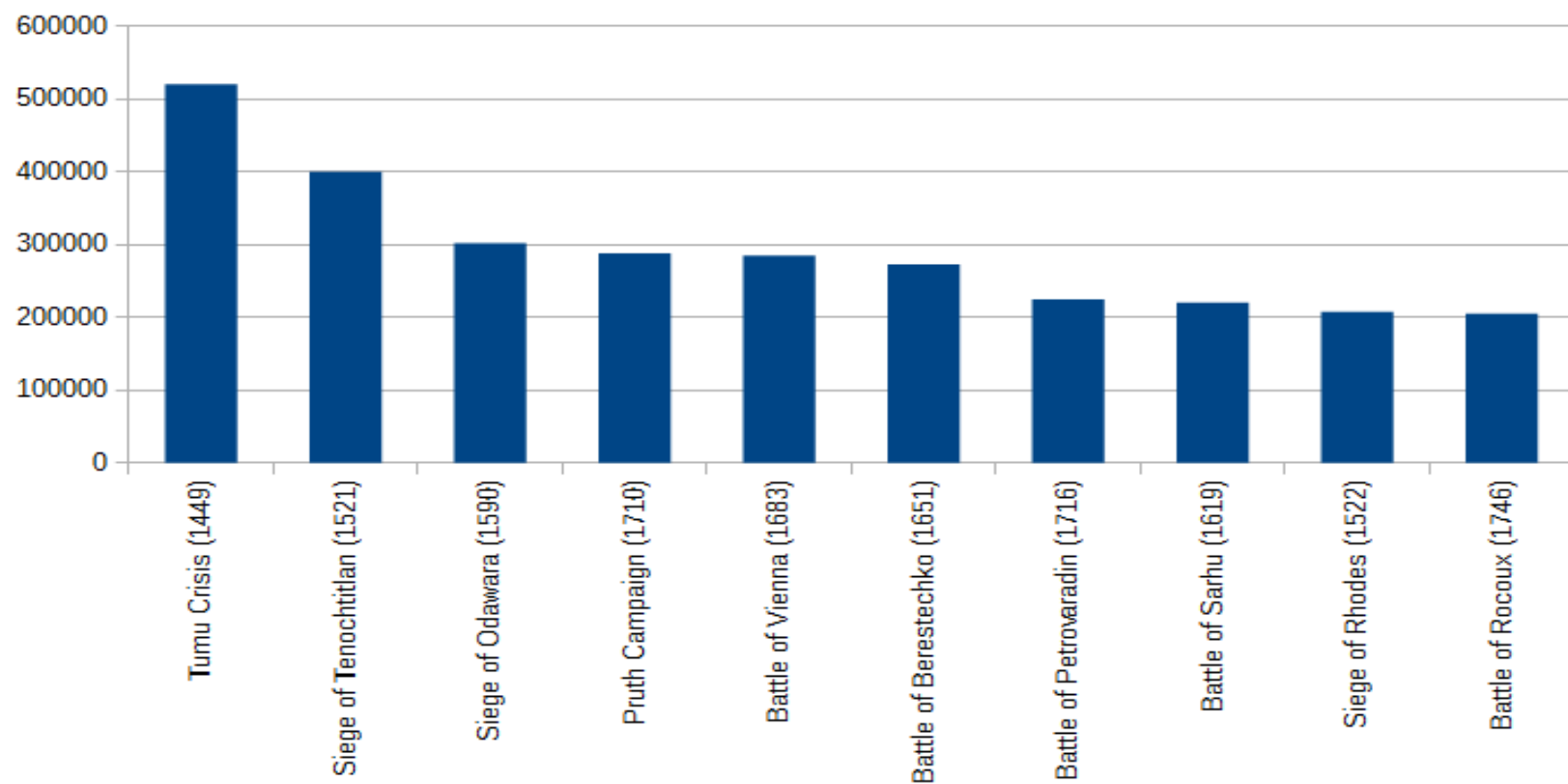
## Wojny z największą sumą sił uczestniczących w bitwach

top 10



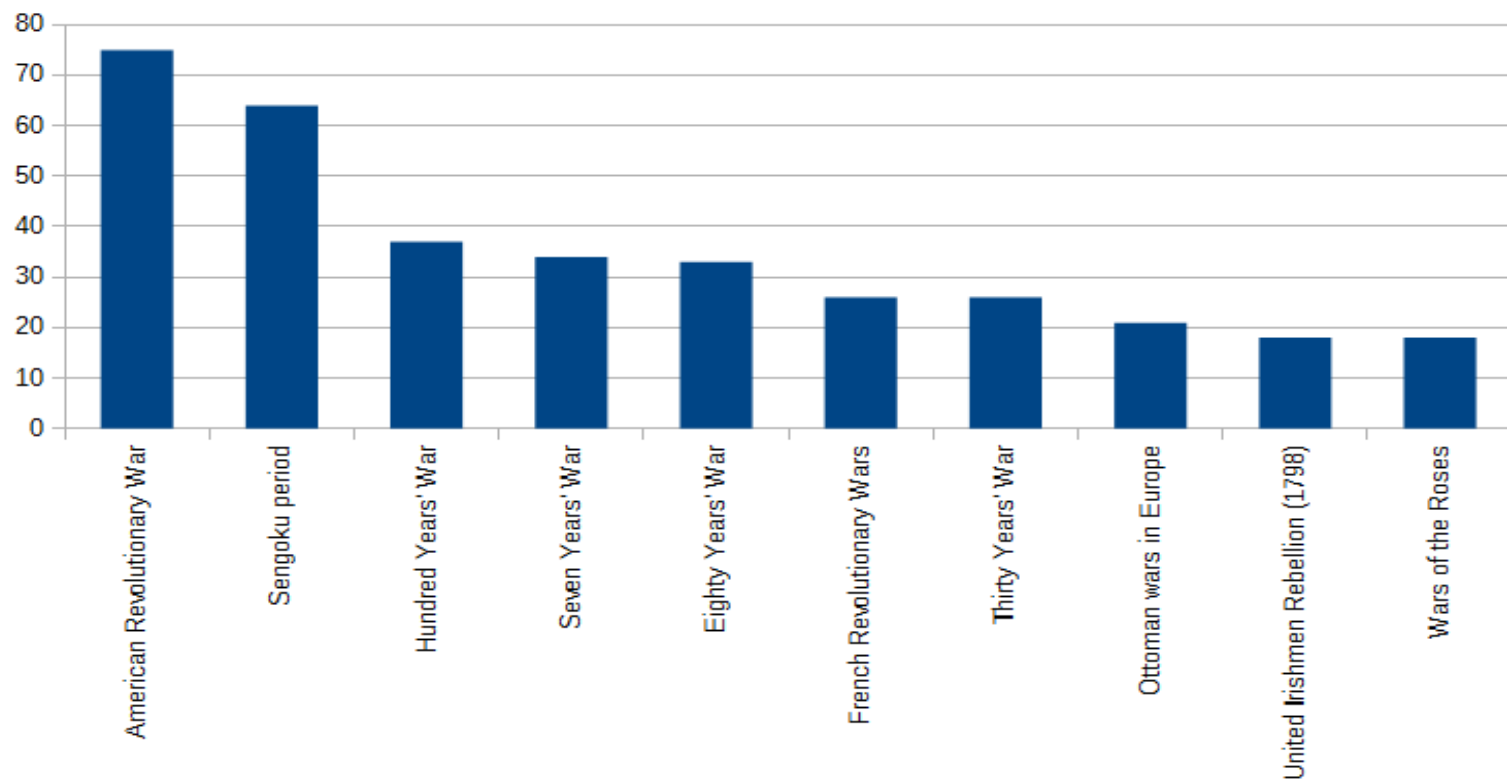
## Suma sił uczestniczących w bitwie

top 10



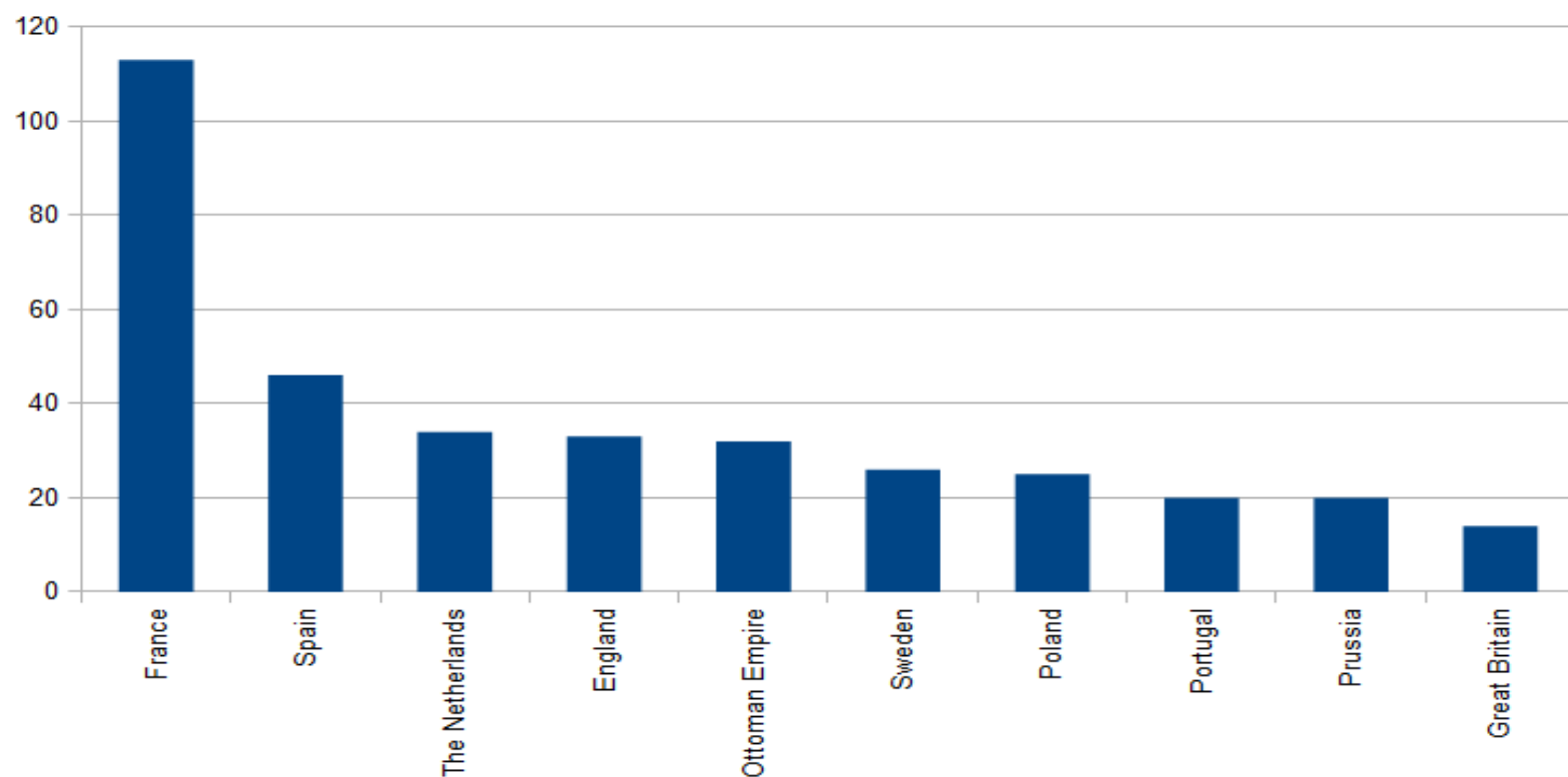
## Wojny o największej liczbie bitew

top 10



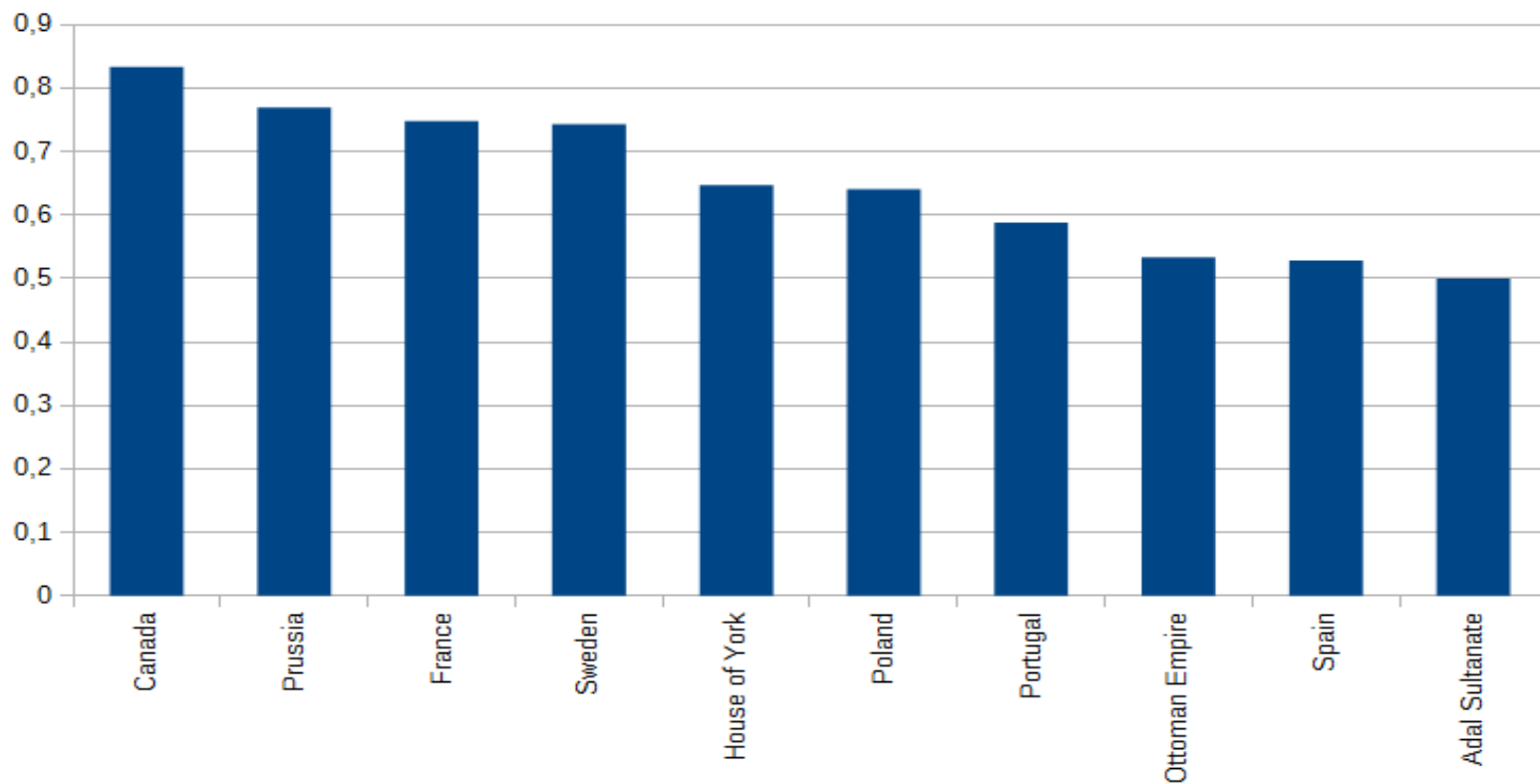
## Strony mające najwięcej zwycięstw w bitwach

top 10



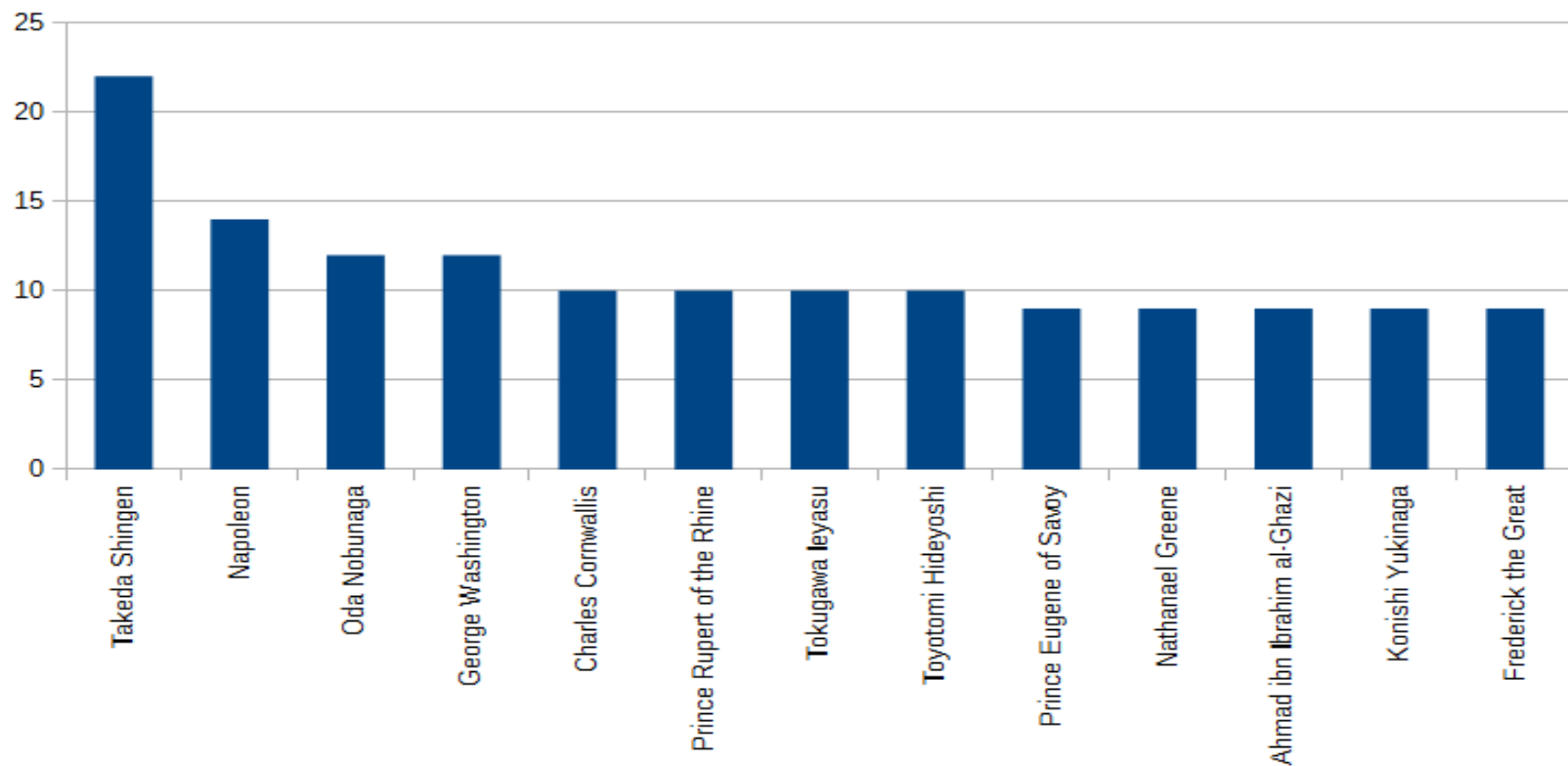
Strony mające największy procent zwycięstw bez uwzględnienia bitew nierozstrzygniętych

top 10

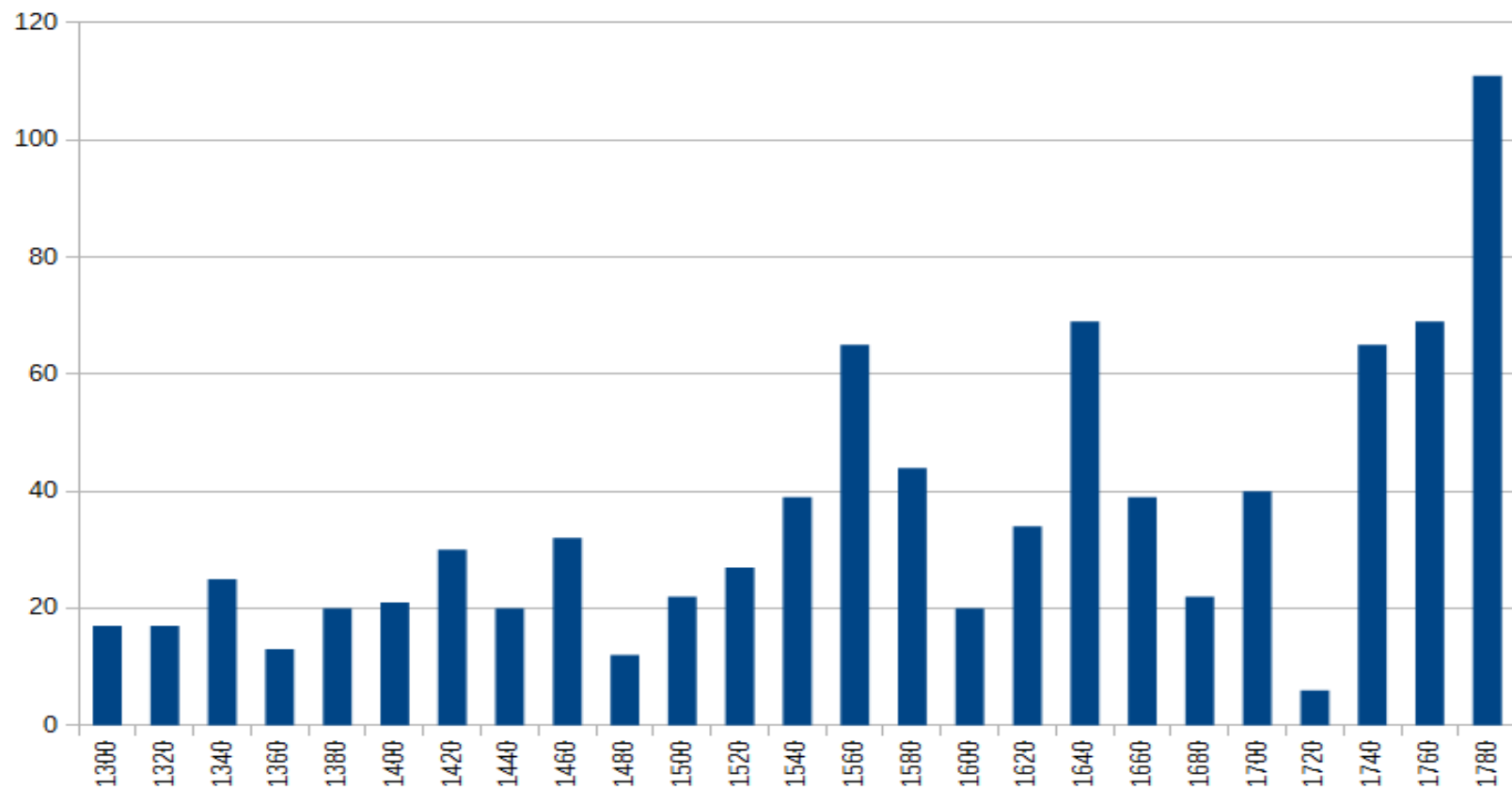


## Dowódcy z największą liczbą stoczonych bitew

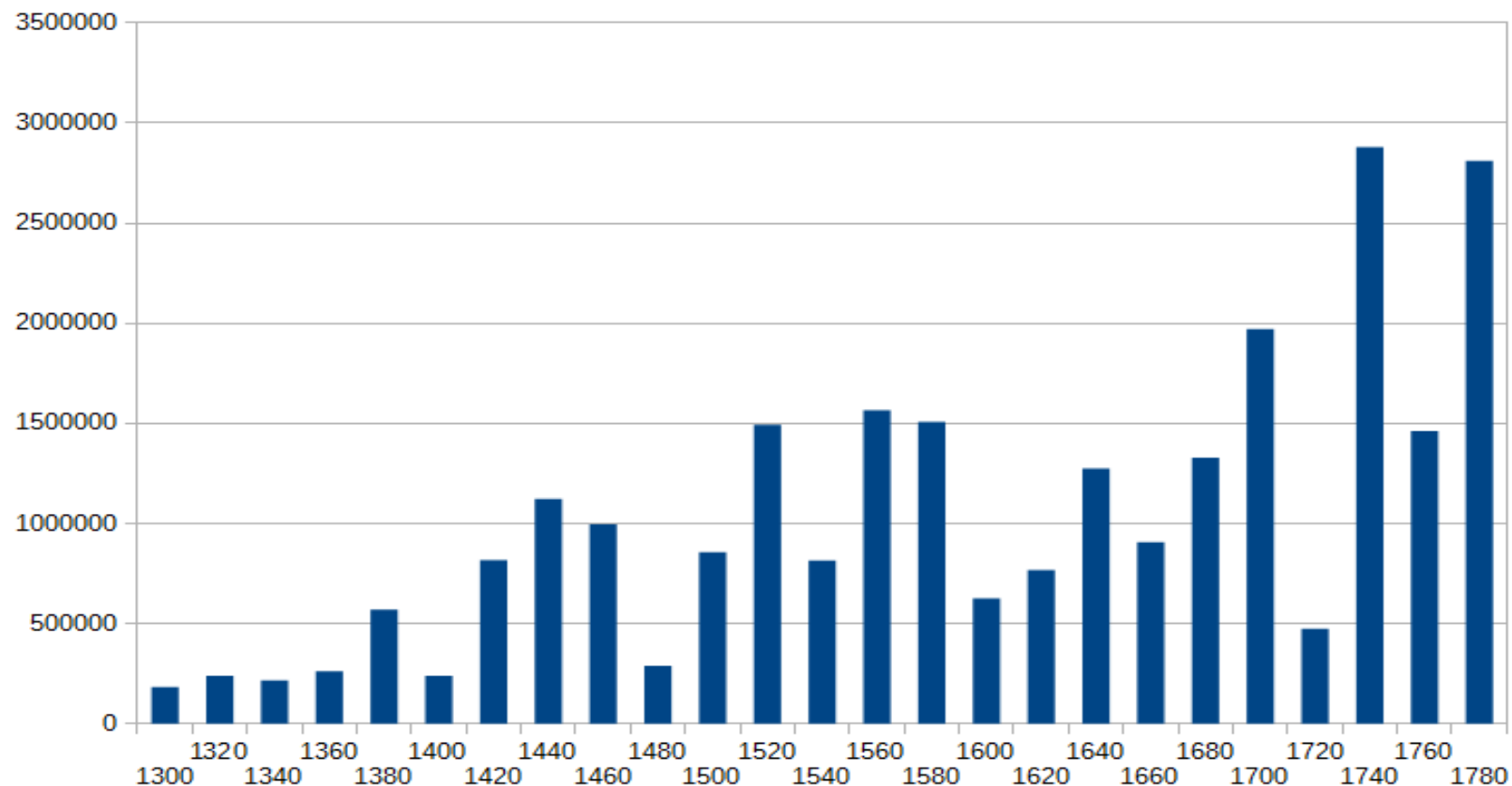
top 10



Liczba bitew w 20-leciach

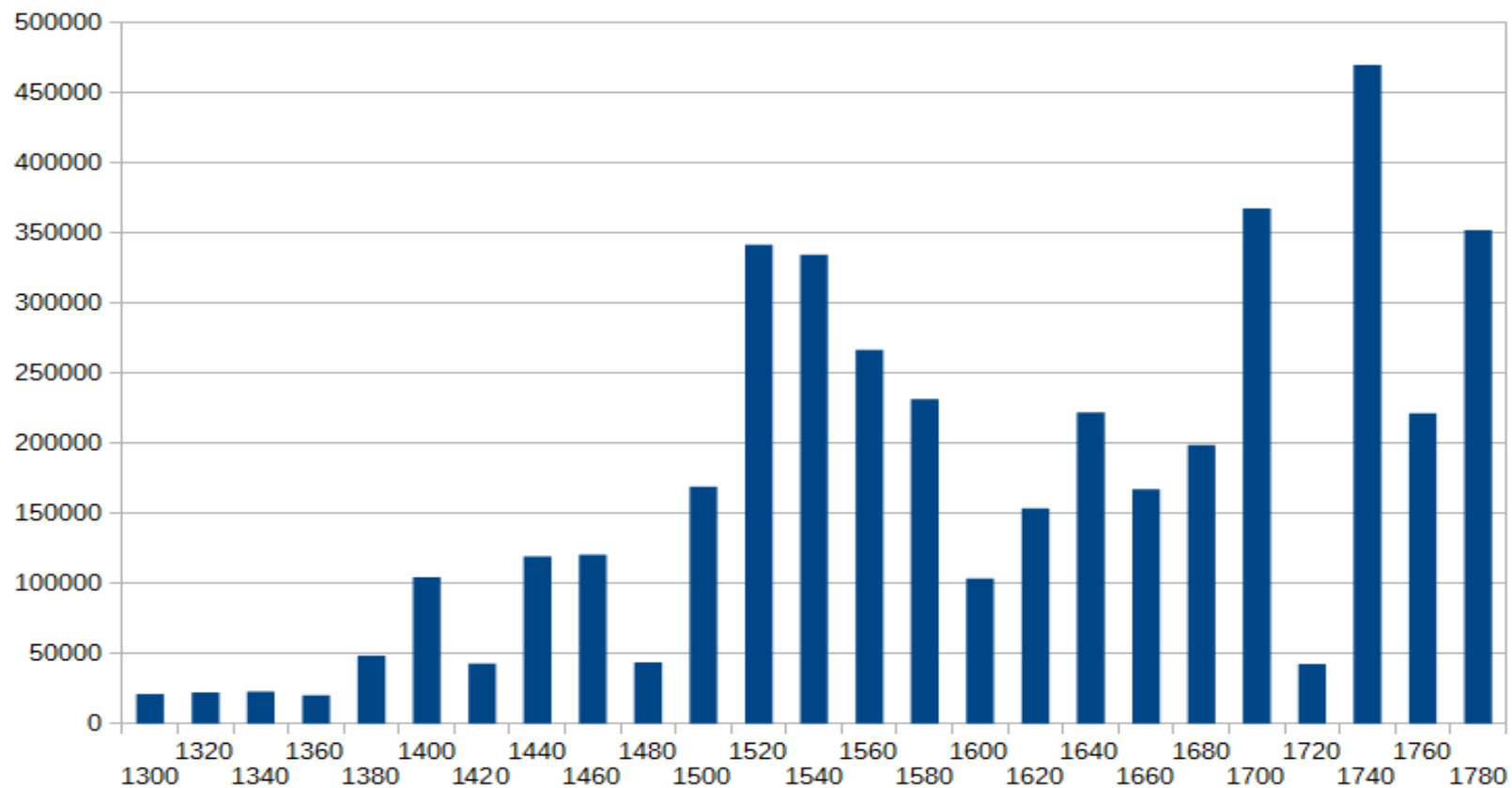


Suma sił w bitwach w 20-leciach

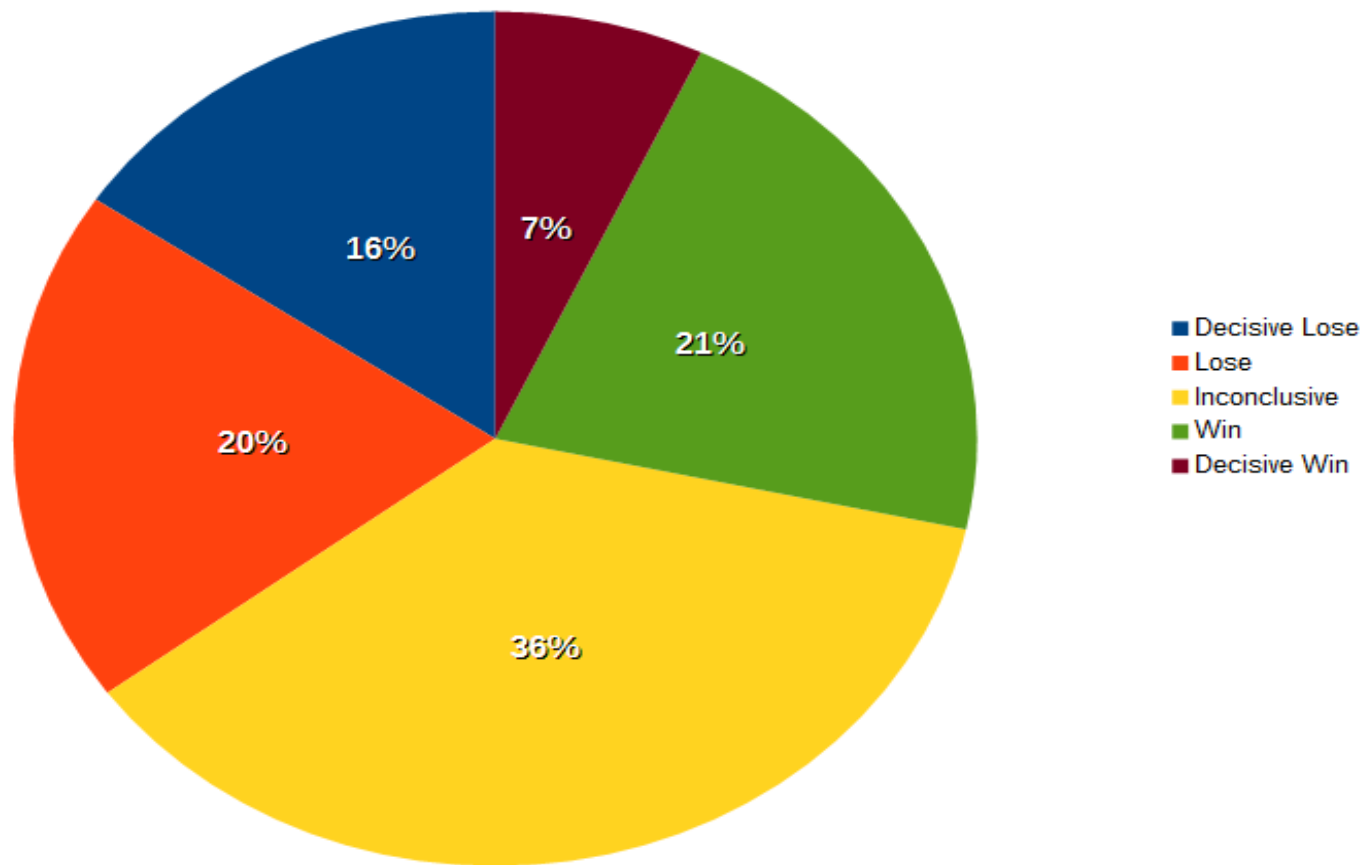


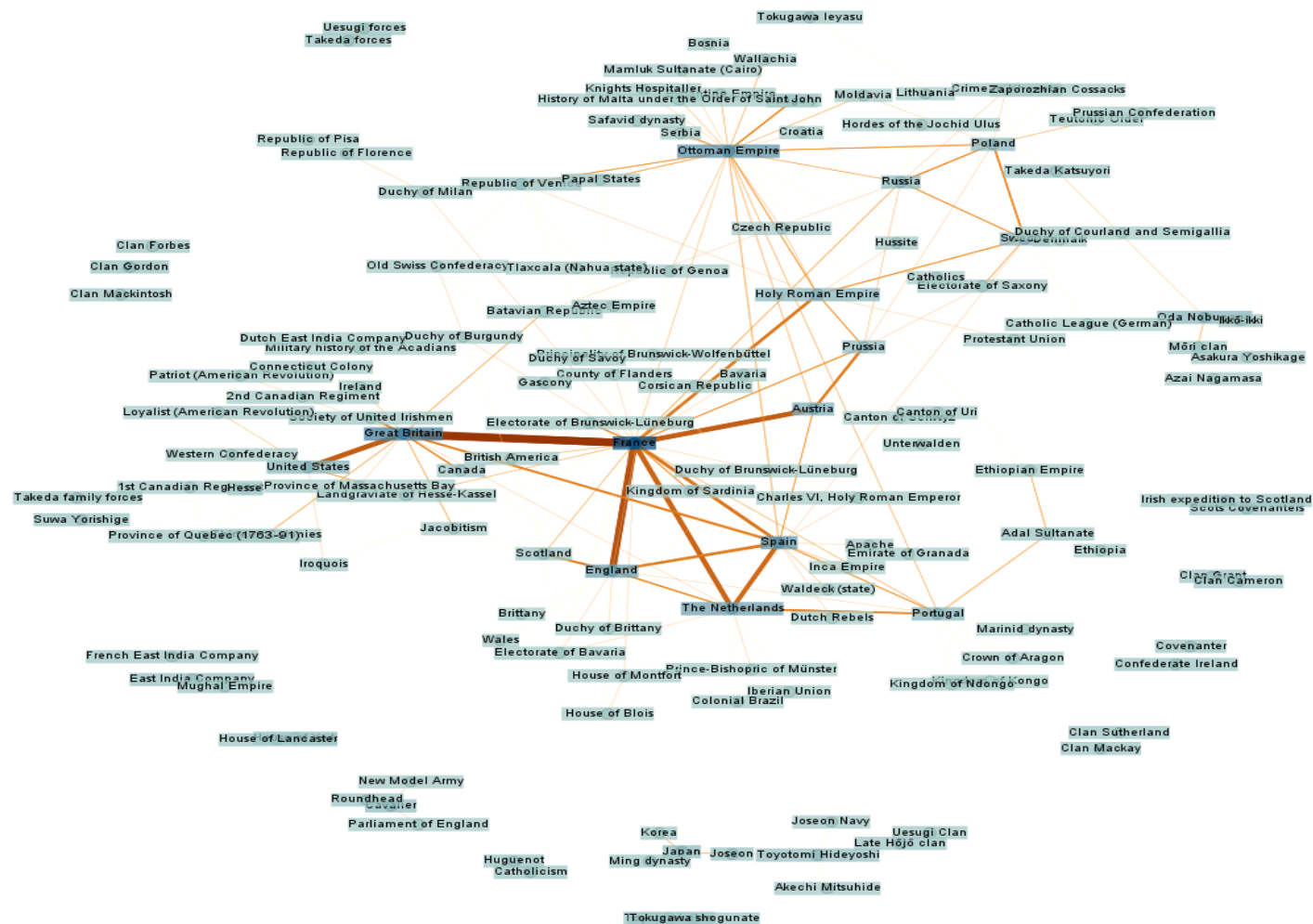


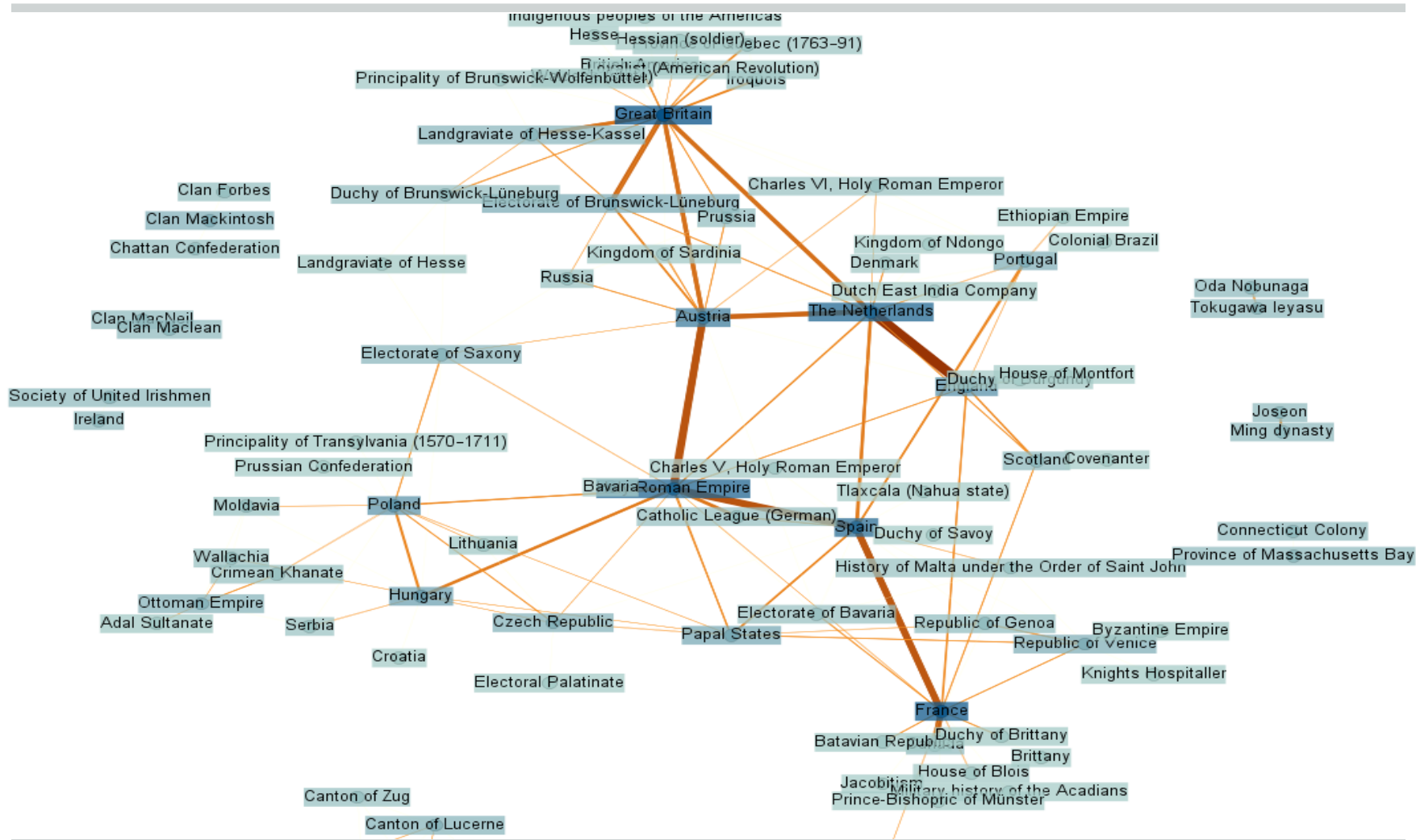
Suma strat w bitwach w 20-leciach

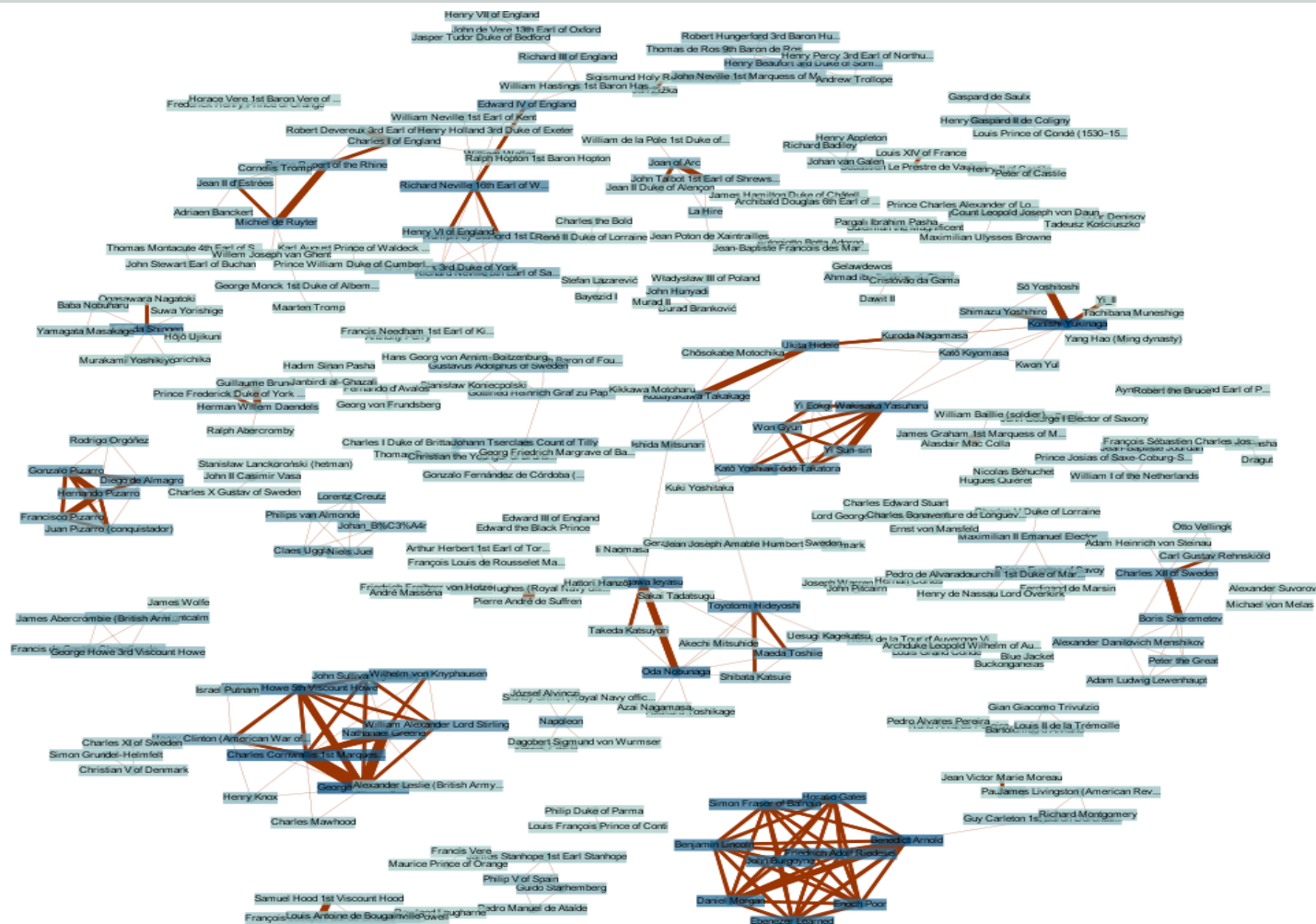


Wynik bitwy z punktu widzenia przeważających sił









# Podsumowanie

---

- Zrealizowano cały plan zawarty w harmonogramie
  - Brak analizy pokoi (niewłaściwa struktura w danych źródłowych)
  - Problemy z przetwarzaniem przez parser - przetwarzanie manualne
-

# **Dziękujemy za uwagę**

---