

Warsztat Badawczy 2

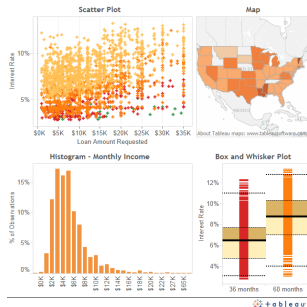
Eksploracyjna analiza danych - przykład

Julian Sienkiewicz

4 marca 2025

Na początkowym etapie analizy danych zaczyna się zwykle od najprostszych metod: jest to **eksploracyjna analiza danych** (*exploratory data analysis - EDA*). Składają się na nią m. in.:

- histogramy,
- wykresy punktowe (*scatterplot*),
- wykresy pudełkowe [skrzynkowe] (*box plot*),

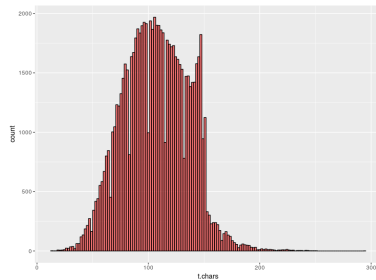
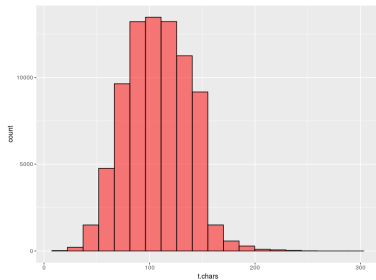


Ogólna definicja Johna Tukeya z lat 60-tych XX w. określa EDA jako

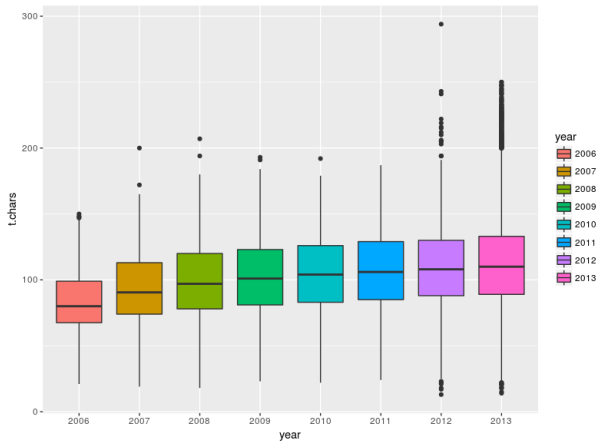
procedury analizowania danych, metody interpretowania wyników takich procedur, sposoby projektowania zbierania danych w celu uczynienia ich analizy łatwiejszą lub bardziej dokładną, oraz wszystkie techniki statystyczne które mogą być stosowane w badaniu danych.

- Sens EDA jest szczególnie widoczny, gdy **nie posiadamy hipotezy badawczej** — często jest tak, że owa hipoteza jest nietrafiona lub też zawęża nasz obszar badań,
- EDA jest, kolokwialnie mówiąc, **grzebaniem się w danych** w celu wytypowania istotnych obserwacji, faktów i cech rozpatrywanego układu,
- zaleca się, aby w trakcie EDA przy korzystaniu ze **statystyk opisowych** odchodzić od średniej i odchylenia na rzecz statystyk odpornościowych (*robust statistics*), takich jak np. mediana,
- Tukey w latach 70-tych XX w. podkreślał, że wywierany jest zbyt duży nacisk na **testowanie hipotez** bezpośrednio na danych,
- postulowane jest wykonanie EDA **w celu** zaproponowania hipotez badawczych

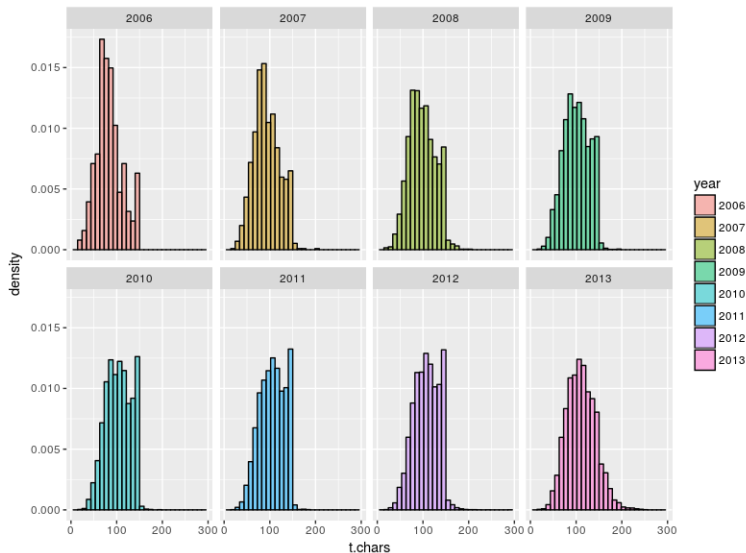
Interesuje nas *t.chars* czyli liczba znaków w tytule



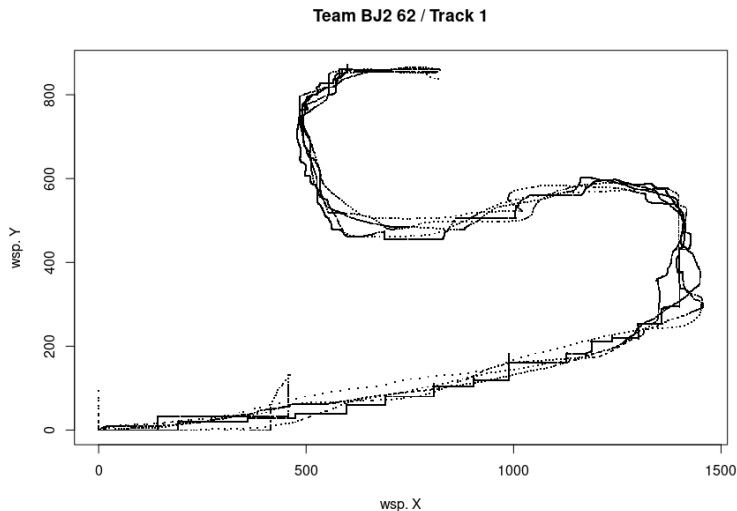
Boxplot *t.chars* na przestrzeni lat



Histogram *t.chars* z rozdziałem na poszczególne lata



Trajektoria zespołu nr. **BJ2 62**, numer toru: **1**



Zmiany współrzędnych zespołu nr. **BJ2 62**, numer toru: **1**

