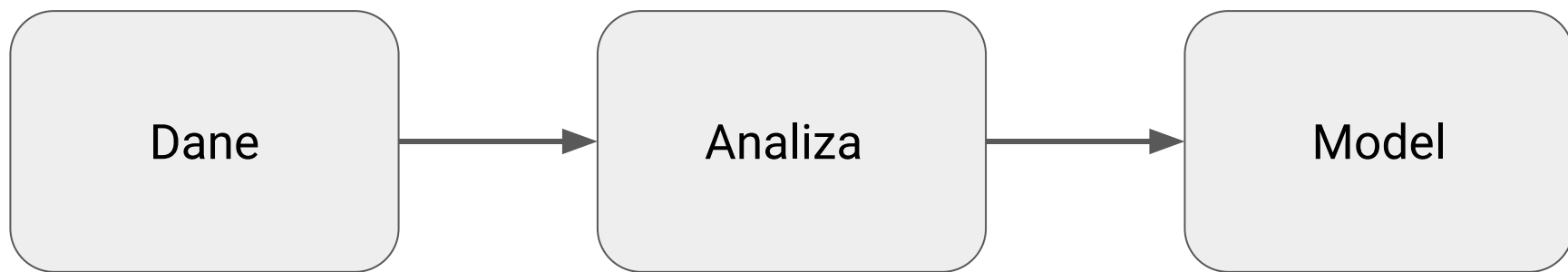


# Analiza EDA przed modelowaniem

EDA, ang. *exploratory data analysis*



Dane

# Dane

a) jakie rozszerzenie, jak wczytać

# Dane

- a) jakie rozszerzenie, jak wczytać
- b) jakie zmienne, jaki wymiar danych

# Dane

- a) jakie rozszerzenie, jak wczytać
- b) jakie zmienne, jaki wymiar danych
- c) zrozumienie co zawierają dane, jaką informację niosą poszczególne kolumny

# Dane

- a) jakie rozszerzenie, jak wczytać
- b) jakie zmienne, jaki wymiar danych
- c) zrozumienie co zawierają dane, jaką informację niosą poszczególne kolumny
- d) czy dane są pełne, ewentualne braki danych



# Problemy?

Co w przypadku “dużej” tabeli?

Analiza

# Analiza

a) ogólne statystyki

# Analiza

a) ogólne statystyki

b) analiza rozkładu jednej zmiennej (zmienne ilościowe i jakościowe)

# Analiza

- a) ogólne statystyki
- b) analiza rozkładu jednej zmiennej (zmienne ilościowe i jakościowe)
- c) korelacje zmiennych

# Analiza

- a) ogólne statystyki
- b) analiza rozkładu jednej zmiennej (zmienne ilościowe i jakościowe)
- c) korelacje zmiennych
- d) analiza dwóch lub więcej zmiennych (na podstawie korelacji zmiennych)

# Analiza

- a) ogólne statystyki
- b) analiza rozkładu jednej zmiennej (zmienne ilościowe i jakościowe)
- c) korelacje zmiennych
- d) analiza dwóch lub więcej zmiennych (na podstawie korelacji zmiennych)
- e) testy statystyczne

Model



# Model

a) podział zbioru względem targetu (stratyfikowany czy nie)

# Model

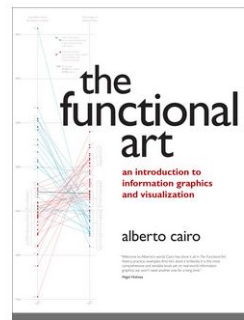
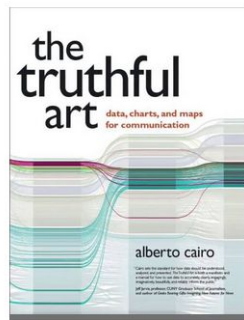
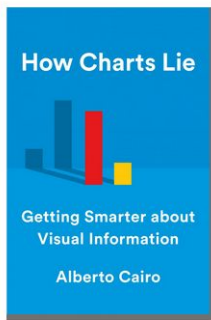
- a) podział zbioru względem targetu (stratyfikowany czy nie)
- b) jaki typ modelu wybrać

Znani w świecie wizualizacji.



## Alberto Cairo

Jest dziennikarzem i projektantem z wieloletnim doświadczeniem w prowadzeniu zespołów graficznych i wizualizacyjnych w wielu krajach.



<http://albertocairo.com/>

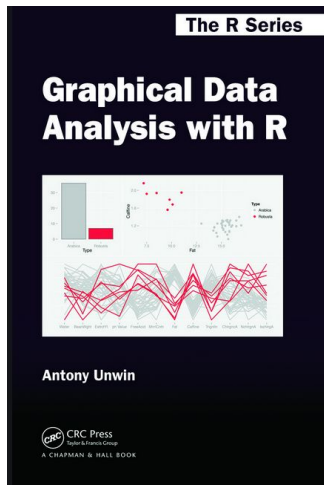
<http://www.thefunctionalart.com/>

<https://www.youtube.com/watch?v=-6stpCiUSWM>



## Antony Unwin

Jest profesorem statystyki zorientowanej komputerowo i analizy danych na Uniwersytecie w Augsburgu w Niemczech oraz członkiem Amerykańskiego Towarzystwa Statystycznego.



Jest współautorem książki "Graphics of Large Datasets" i współredaktorem "Handbook of Data Visualization".

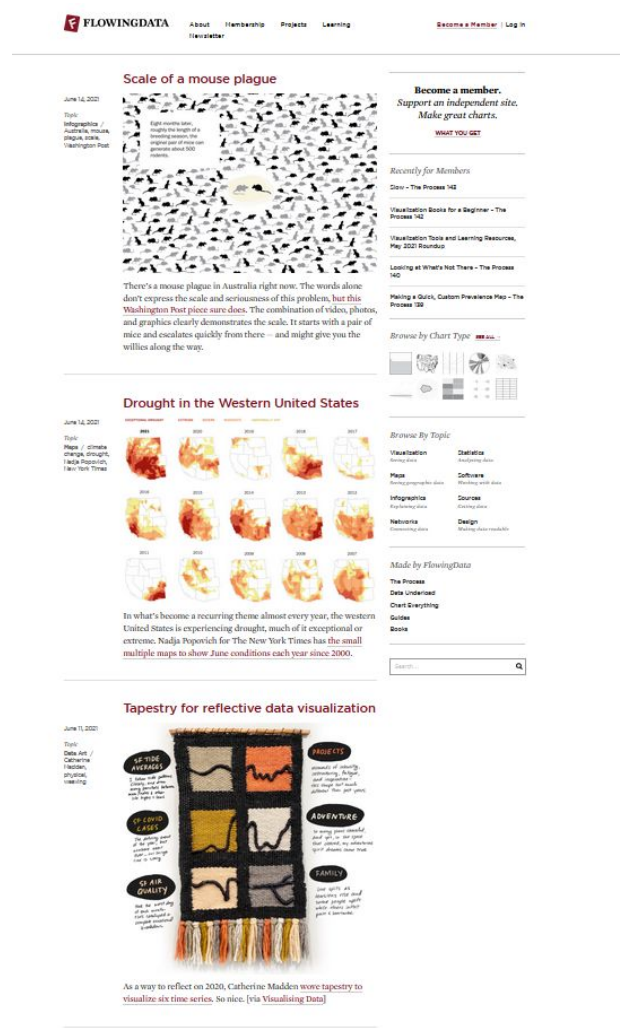
<http://www.gradaanwr.net/author/>

# Blog o tematyce wizualizacji

Nathan Yau

FlowingData bada, jak wykorzystujemy analizę i wizualizację do zrozumienia danych i nas samych.

<https://flowingdata.com/>



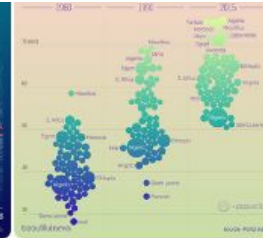
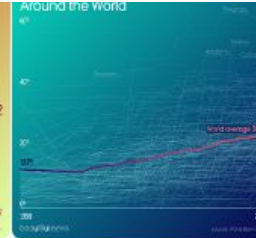
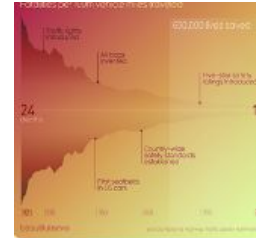
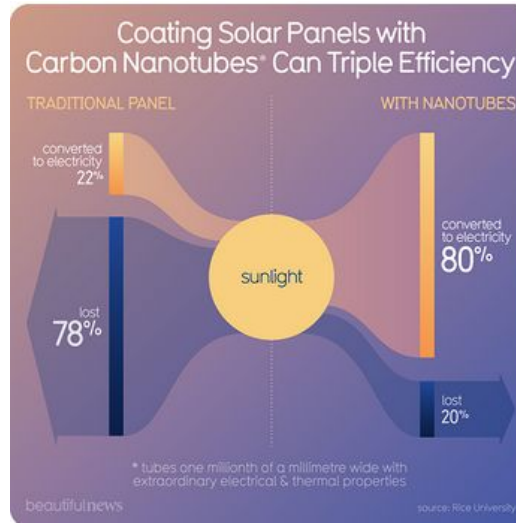


## Hanna Piotrowska (Drycz)

Projektantka graficzna, skupiająca się głównie na wizualizacji danych, brandingu i projektowaniu książek, z silnym zainteresowaniem naukami o danych, badaniami percepcji i edukacją.

<https://www.behance.net/hannapio>

**Beautiful News** is a collection of good news, uplifting statistics and facts. Inspired by the Hans Rosling's idea of Factfulness, Beautiful News releases a chart every day to move our attention beyond dramatic news headlines to the positive trends and slow developments that go unseen, uncelebrated.



<https://www.behance.net/gallery/96978161/Beautiful-News-infographics-data-visualizations>  
<https://informationisbeautiful.net/>

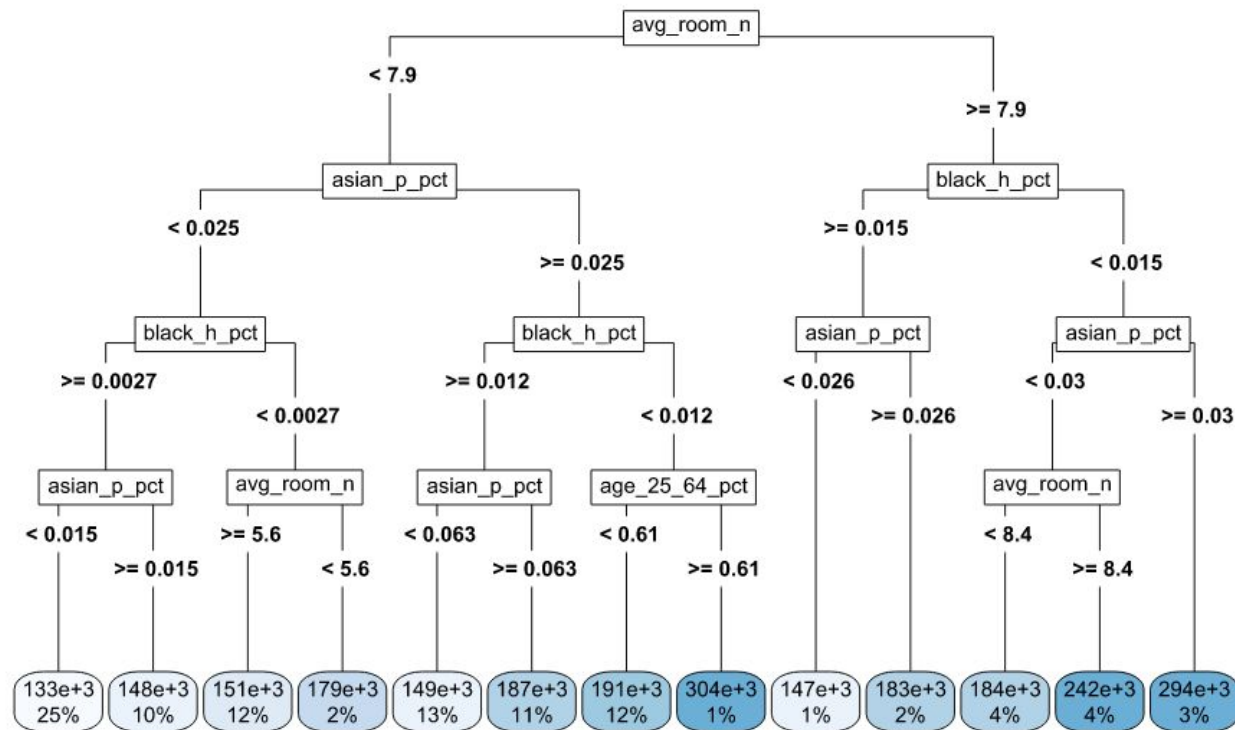


# Wizualizacje modeli

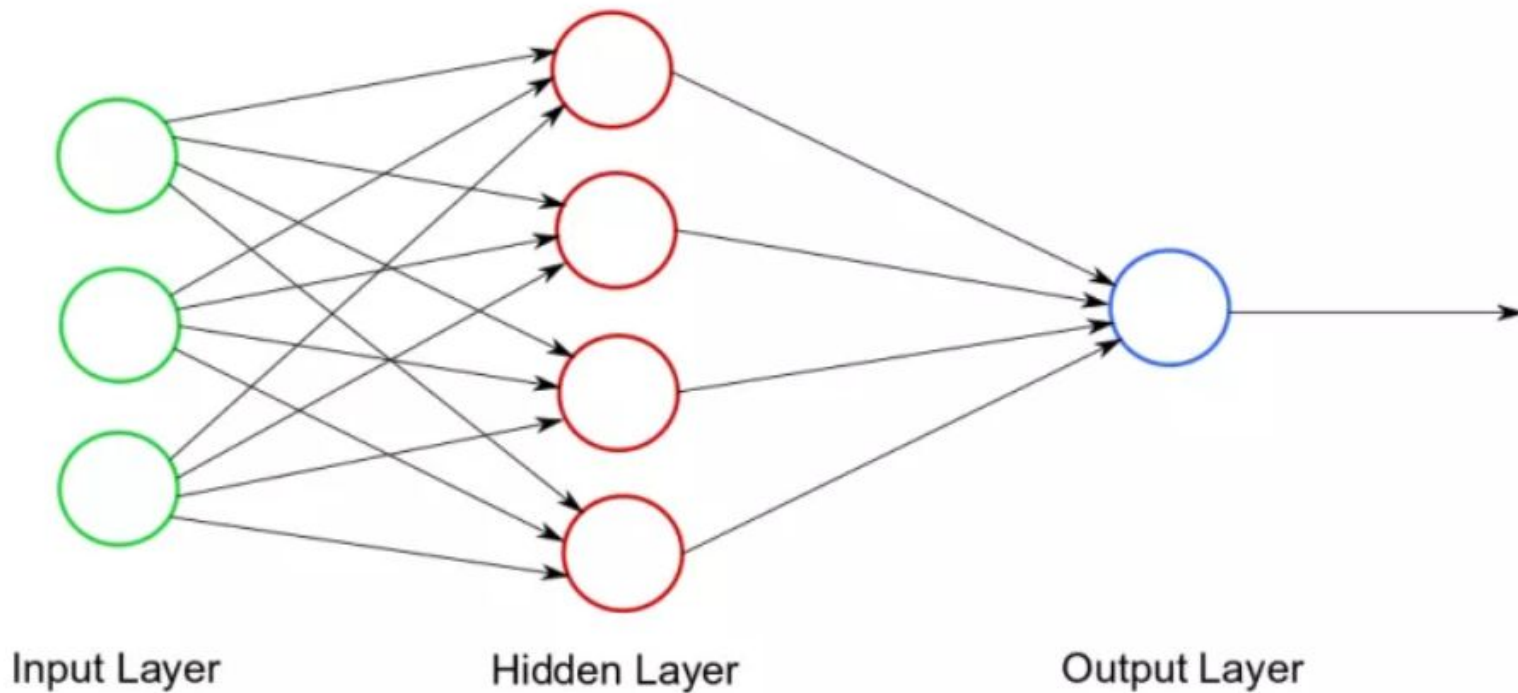
# W jaki sposób możemy wizualizować modele?

- pokazując ich strukturę

# Drzewo decyzyjne

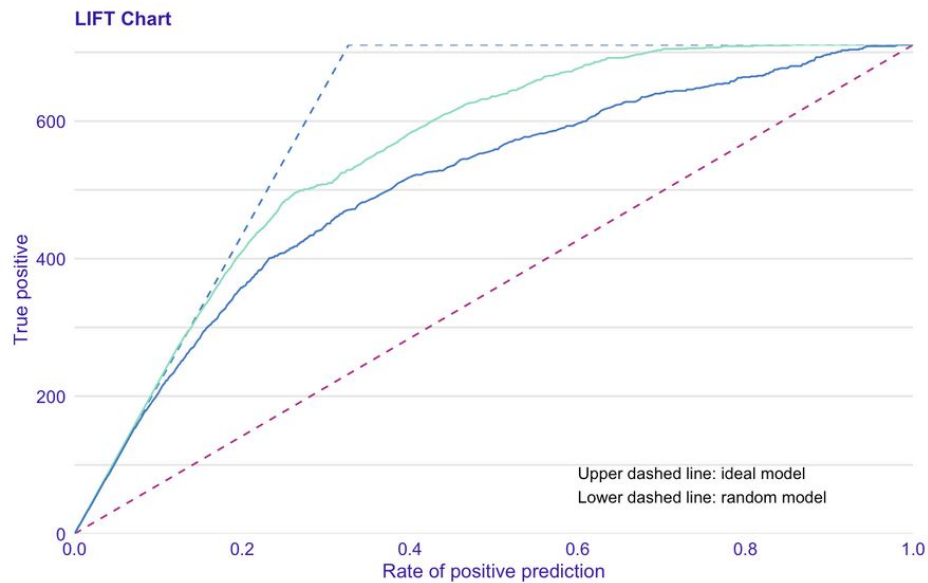
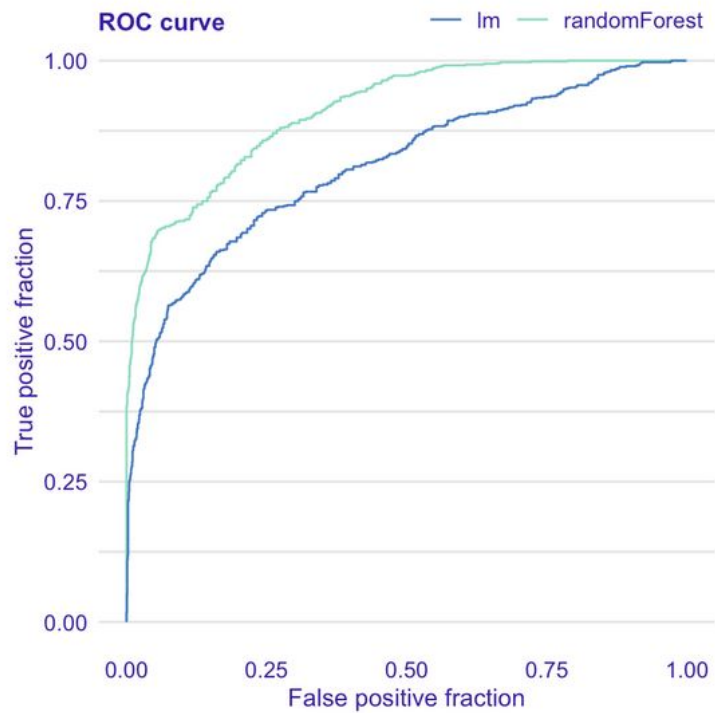


# Sieć neuronowa

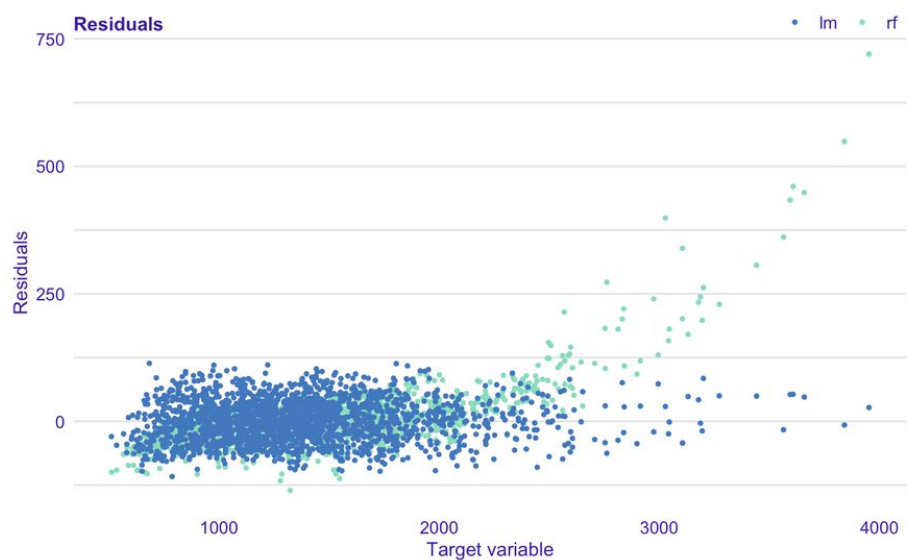
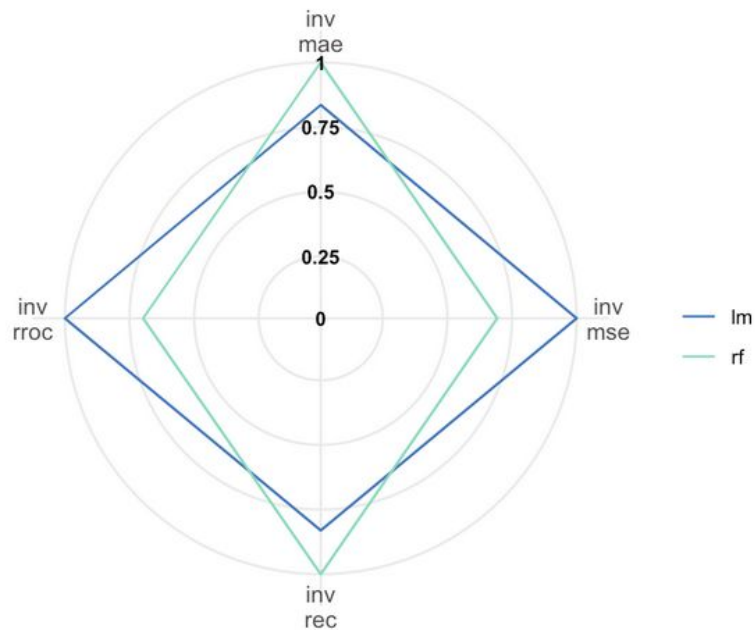


# W jaki sposób możemy wizualizować modele?

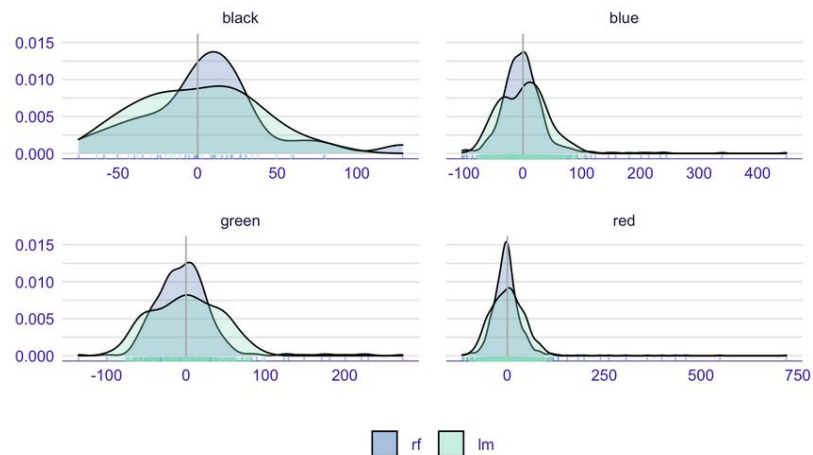
- pokazując ich skuteczność (miara)



## Model ranking radar



## Residuals density by colour

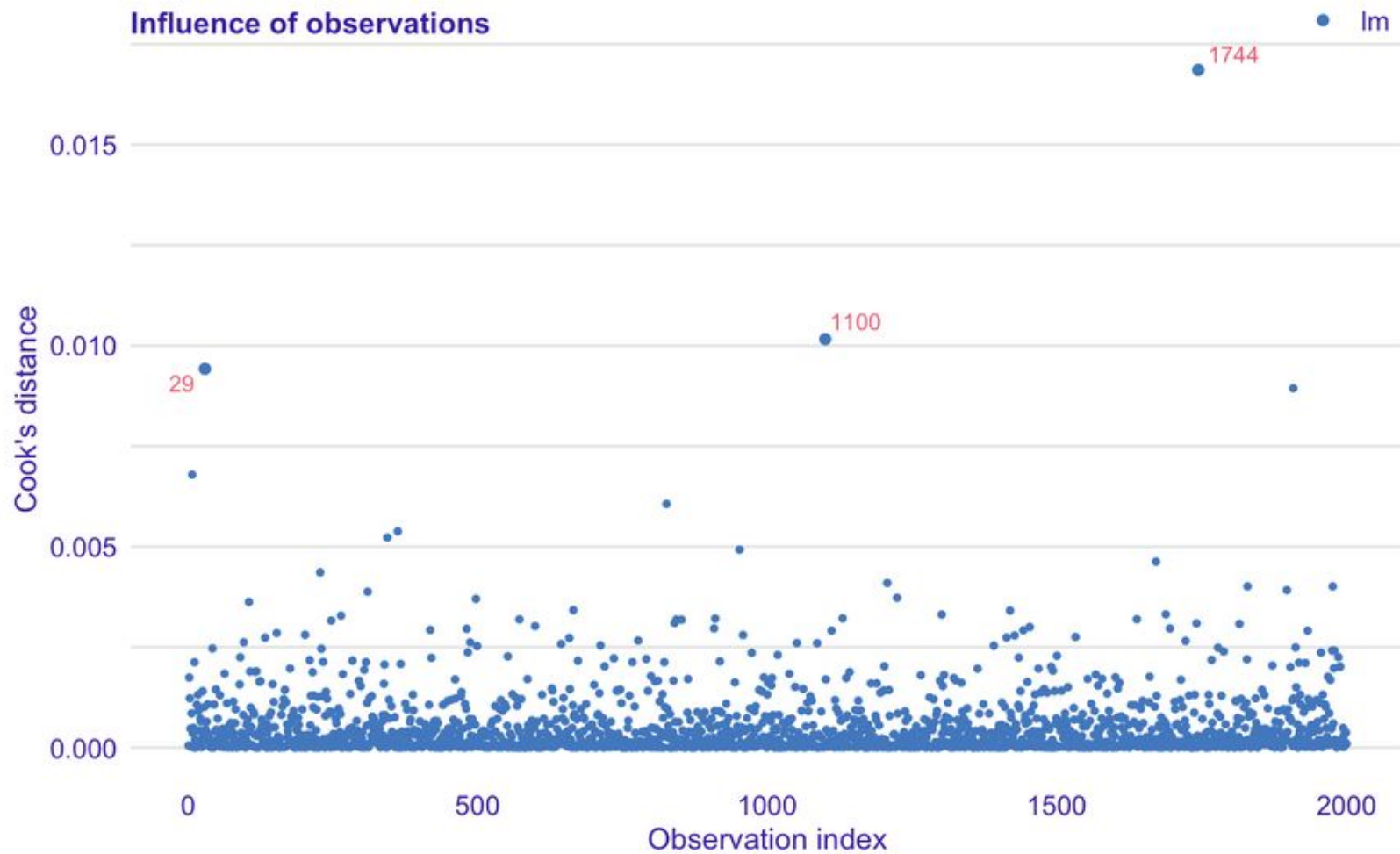


# W jaki sposób możemy wizualizować modele?

- analizując wyniki dla obserwacji

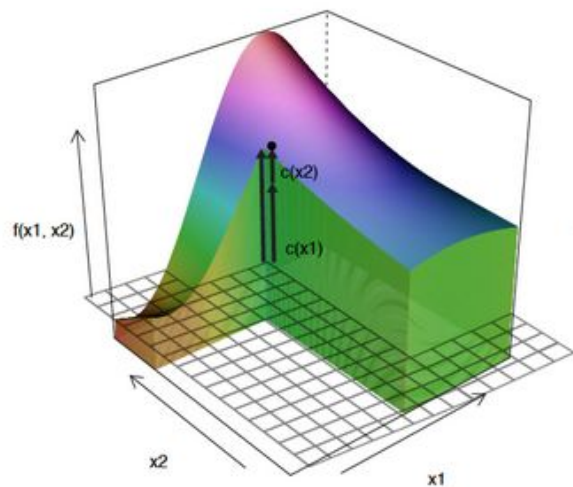


## Influence of observations

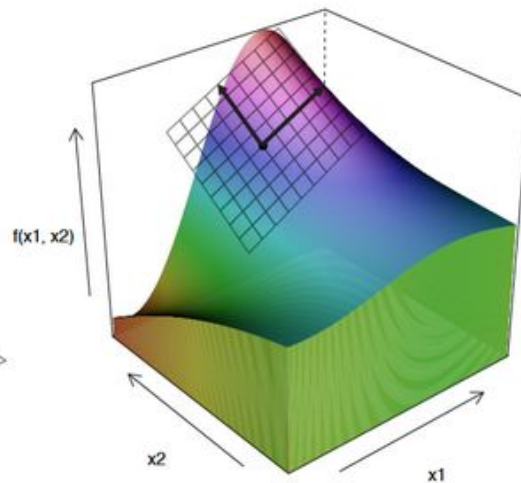


# Wyjaśnienia lokalne

A)



B)



C)

