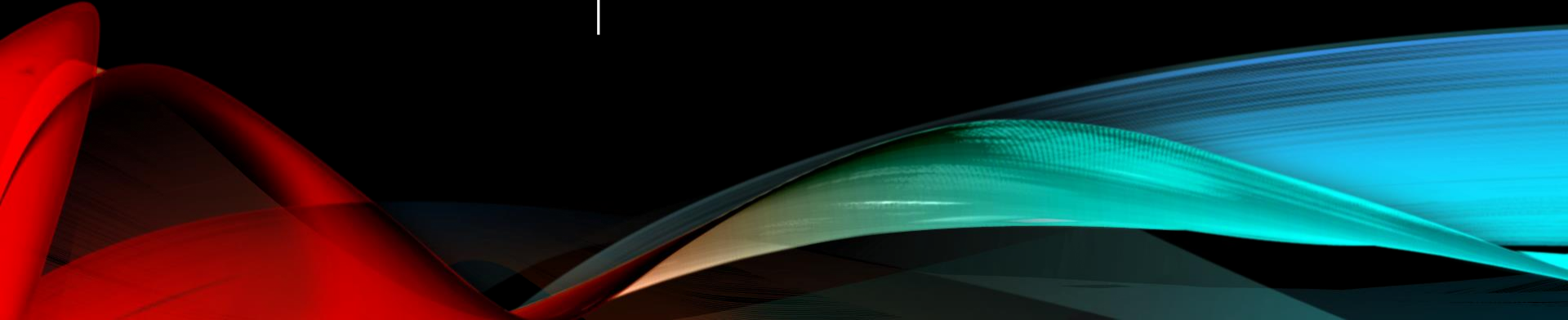


Anna Kozak
Magda Prokopczuk

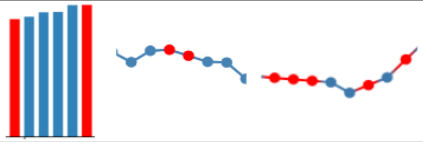
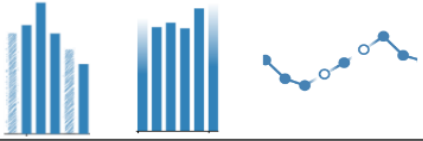
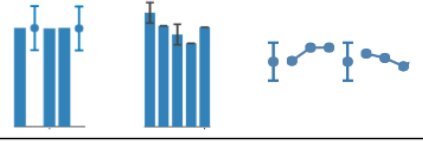

WHERE'S MY DATA? EVALUATING VISUALIZATIONS WITH MISSING DATA



CEL BADANIA

Zmierzenie, w jaki sposób techniki imputacji i wizualizacji danych wpływają na postrzeganie, jakość i dokładność wizualizacji z brakami danych.

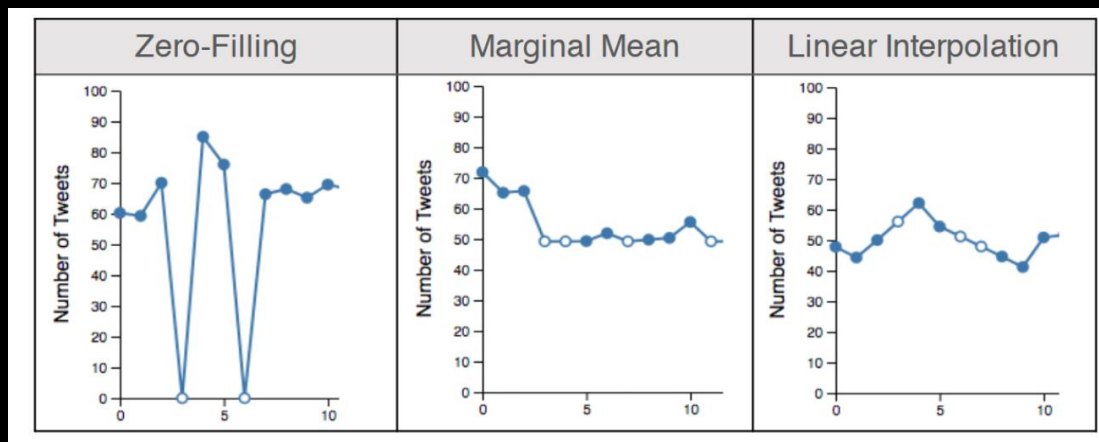
CZTERY KATEGORIE WIZUALIZACJI DO UKAZANIA BRAKÓW DANYCH

Highlight	
Downplay	
Annotation	
Information Removal	

- Highlight - wykorzystanie koloru do zwrócenia uwagi na brakujące punkty danych
- Downplay - zmniejszenie zaznaczenia wartości w stosunku do pozostałych
- Annotation - dodanie informacji statystycznych, słupki błędów
- Information removal - usuwanie brakujących wartości

METODY INPUTACJI DANYCH

- zastąpienie wartości brakujących przez zero
- interpolacja liniowa pomiędzy sąsiednimi dostępnymi obserwacjami
- zastąpienie przez wartość średnią



POSTAWIONE HIPOTEZY

01

Postrzegana jakość danych i dokładność zmniejszy się wraz ze wzrostem liczby brakujących danych.

02

Wyróżnianie metod generuje wyższą postrzeganą jakość danych niż metody pomniejszania i usuwania informacji.

03

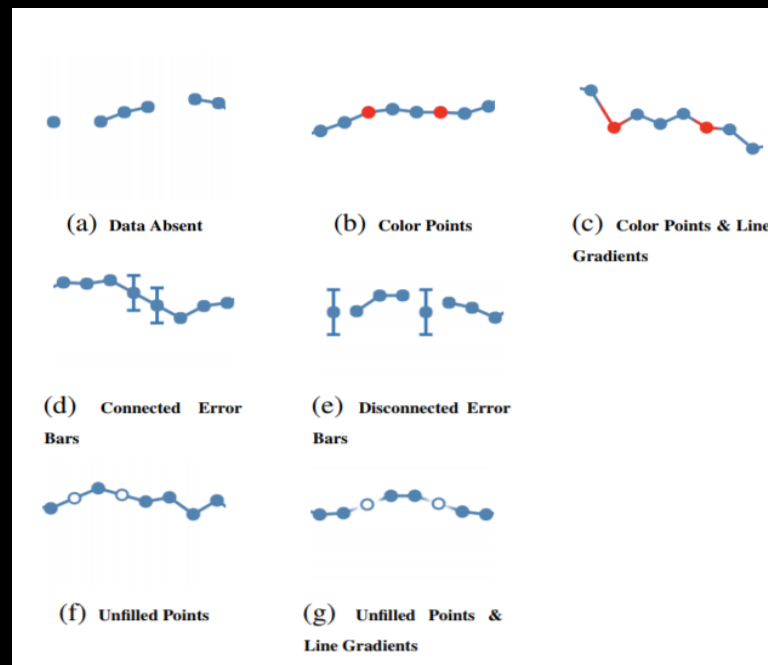
Interpolacja liniowa doprowadzi do wyższej postrzeganej jakości danych niż średnia lub wypełnienie wartością zero.

04

Przypisane wartości powodują wyższą postrzeganą jakość danych niż wartości usunięte.

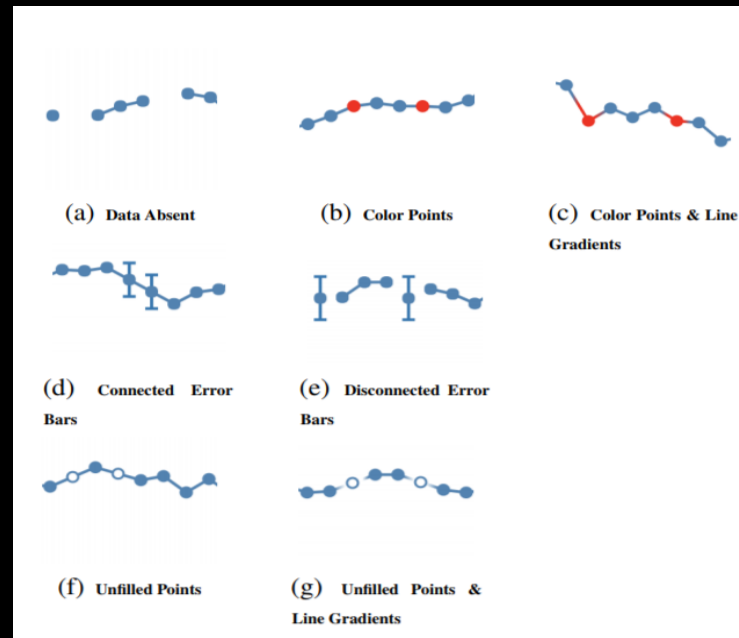
WYKRES LINIOWY

- Uśrednianie:
 - Połączone słupki błędów i kolorowe punkty prowadzą do wyższej postrzeganej jakości danych, podczas gdy braki danych i niepołączone słupki do niższej.
 - Przybliżenie liniowe okazuje się dokładniejsze niż uśrednienie lub imputacja zerem.
 - Połączone słupki błędów wraz z imputacją zerem prowadzą do znacząco wyższej postrzeganej jakości danych niż jakikolwiek inny sposób wizualizacji połączony z imputacją zerem.

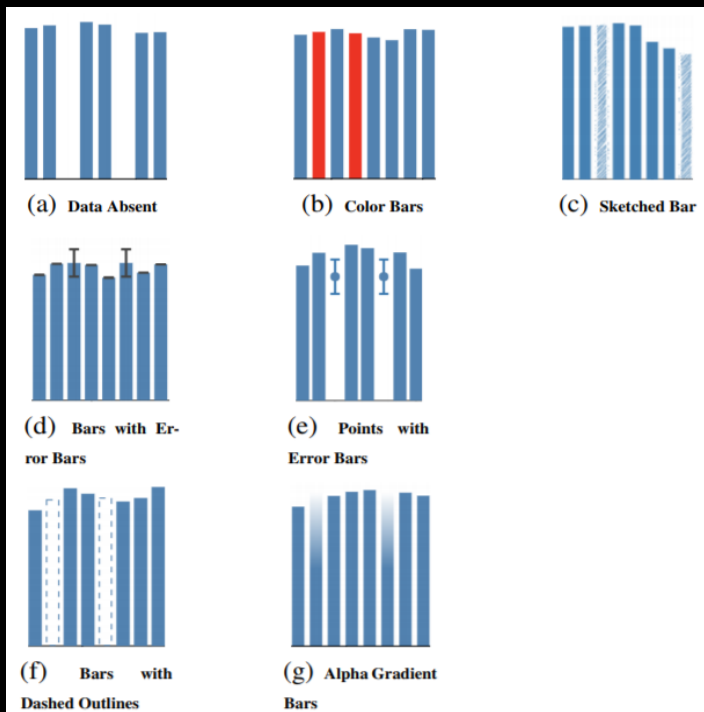


WYKRES LINIOWY

- Detekcja trendu:
 - Połączone słupki błędów i kolorowe punkty prowadzą do wyższej postrzeganej jakości danych, podczas gdy braki danych do niższej.
 - Wpływ niepołączonych słupków błędów na detekcję trendu znacząco różni się od wpływu na uśrednianie.
 - Przybliżenie liniowe okazuje się dokładniejsze niż uśrednienie lub imputacja zerem.
 - Połączone słupki błędów wraz z imputacją zerem prowadzą do znacząco wyższej postrzeganej jakości danych niż jakikolwiek inny sposób wizualizacji połączony z imputacją zerem.



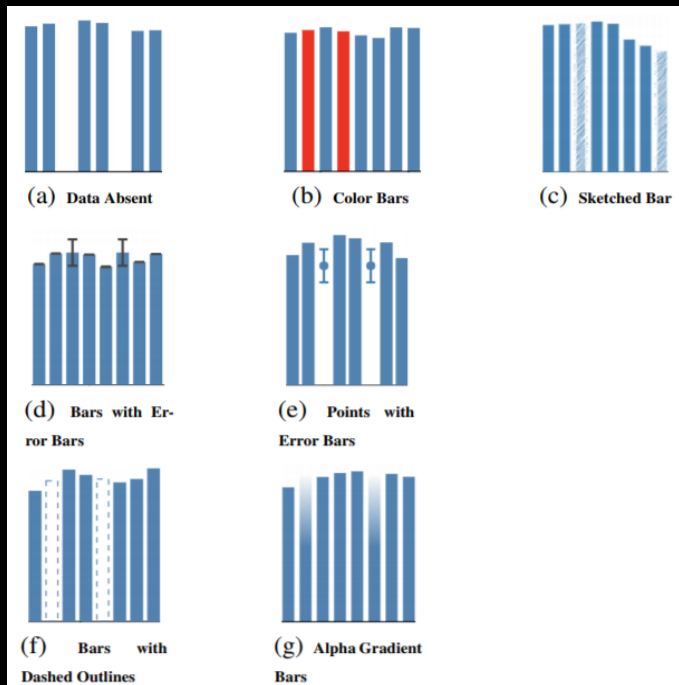
WYKRES SŁUPKOWY



- Uśrednianie:

- Słupki z zaznaczonymi błędami, zaznaczone kolorem oraz gradientem prowadzą do wyższej postrzeganej jakości danych, podczas gdy punkty z słupkami błędów oraz techniki zmniejszania zaznaczenia do niższej postrzeganej jakości danych.
- Przybliżenie liniowe nie różni się znacząco w oczach odbiorcy wykresu od uśrednienia, oba okazują się dokładniejsze niż imputacja zerem.
- Słupki z zaznaczonymi błędami wraz z interpolacją liniową są bardziej odporne na ilość braków danych niż punkty z błędami.

WYKRES SŁUPKOWY



- Detekcja trendu:
 - Słupki z zaznaczonymi błędami prowadzą do wysokiej postrzeganej jakości danych. Braki danych, linia przerywana oraz punkty z błędami skutkowały niższą jakością wizualizacji.
 - Przybliżenie liniowe prowadzi do najwyższej jakości wizualizacji, różniąc się znacząco w oczach odbiorcy wykresu od uśrednienia, które okazuje się dokładniejsze niż imputacja zerem.



PODSUMOWANIE

- Postrzegana jakość danych maleje wraz ze wzrostem liczby braków danych.
- Podkreślenie braków danych odbierane jest lepiej niż zmniejszanie ich widoczności lub całkowity brak.
- Wizualizacja wykorzystująca aproksymację liniową braków danych rejestrowana jest jako mająca wyższą jakość niż pozostałe wymienione metody.
- Wizualizacja łamiącą ciągłą strukturę redukuje postrzeganą jakość danych.

BIBLIOGRAFIA

Hayeong Song, Danielle Albers Szafir, Where's my data? Evaluating Visualizations with Missing Data, IEEE VIS 2018

http://cmci.colorado.edu/visualab/papers/song_VIS_2018.pdf