

# Techniki Wizualizacji Danych – praca domowa 2

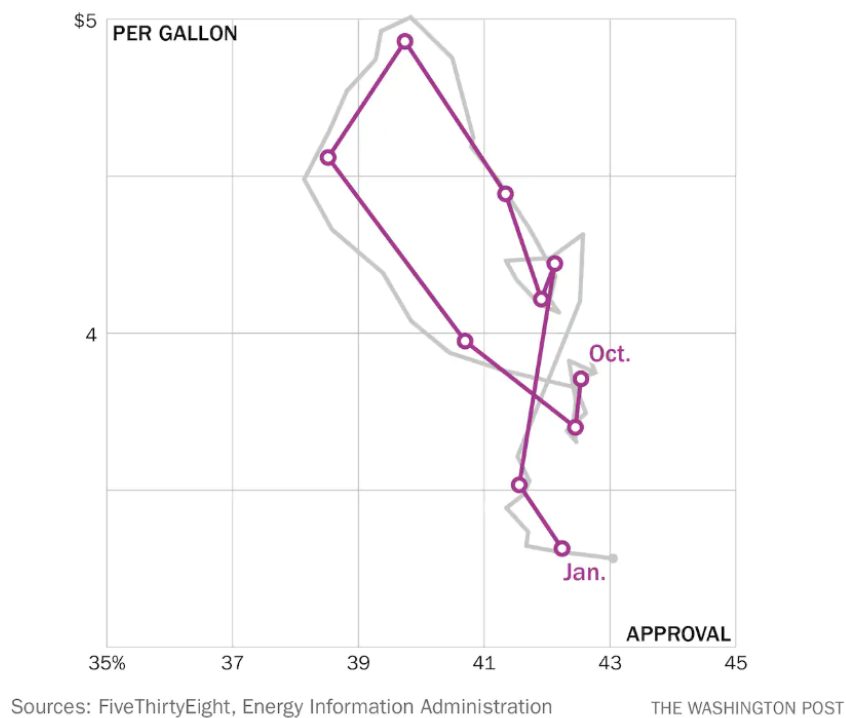
Jeremi Kurdek

listopad 2022

## 1 Poprawiana wizualizacja

### Gas prices vs. Biden approval

From the beginning of the year to mid-October.



Rysunek 1: Wizualizacja poparcia prezydenta Bidena vs ceny paliwa

Wizualizacja wybrana w celu poprawy 1 została znaleziona na stronie gazety Washington Post.

<https://www.washingtonpost.com/politics/2022/10/18/gas-prices-election-biden/>

Wykres przedstawia poparcie dla urzędującego prezydenta Stanów Zjednoczonych oraz ceny paliw w tymże kraju w zależności od czasu. Obserwacje zostały uśrednione dla poszczególnych miesięcy.

## 2 Dane

Dane odczytane z wykresu zostały umieszczone w tabeli 1.

Tabela 1: Dane odczytane z wykresu

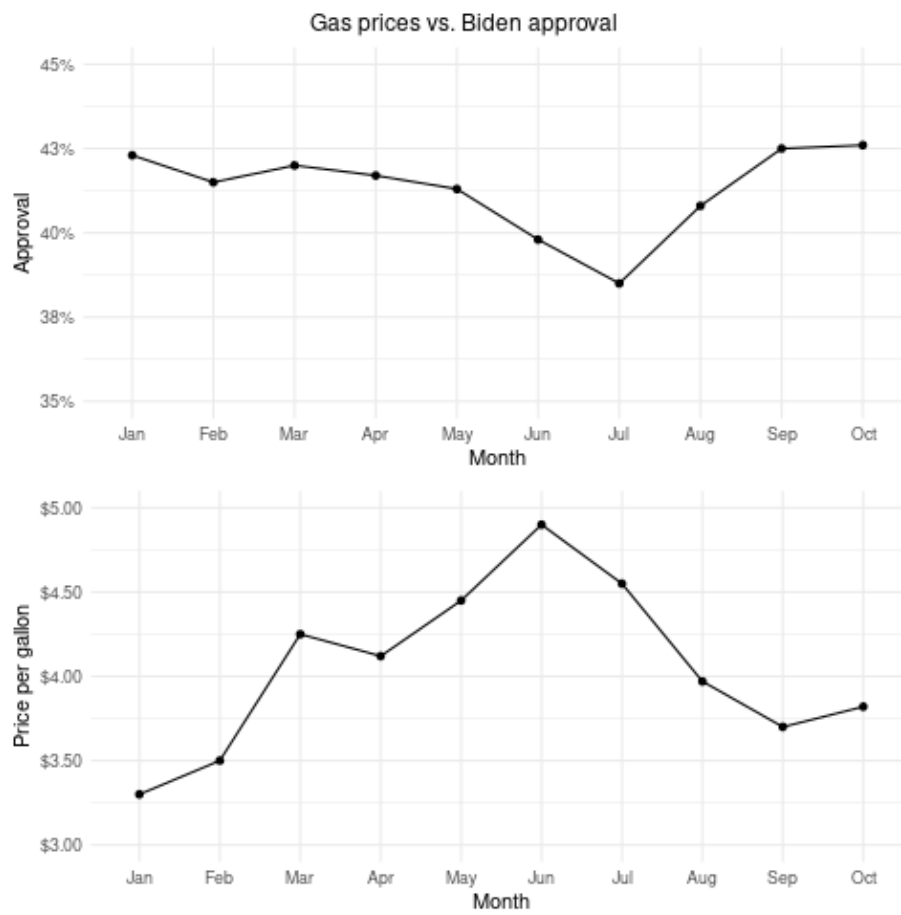
Month	Approval	Price per gallon (\$)
Jan	0.423	3.3
Feb	0.415	3.5
Mar	0.42	4.25
Apr	0.417	4.12
May	0.413	4.45
Jun	0.398	4.9
Jul	0.385	4.55
Aug	0.408	3.97
Sep	0.425	3.7
Oct	0.426	3.82

## 3 Wizualizacje po poprawkach

Zostały przygotowane dwa warianty poprawek do zadanego wykresu. Pierwszy wariant obrazuje rozbicie wykresu na dwa prostsze, łatwiejsze w interpretacji dla niewprawionego odbiorcy. Drugi zachowuje oryginalny typ wykresu, wprowadza jednak poprawki zwiększające czytelność.

### 3.1 Dwa wykresy

Wykres 2 prezentuje dane rozbite na wykresy liniowe.



Rysunek 2: Dane rozbite na 2 wykresy

Kod użyty do przygotowania wizualizacji:

```
library(ggplot2)
library(dplyr)
library(gridExtra)

df <- read.csv("biden.csv")

df %>%
  mutate(Month = factor(Month, levels = Month)) %>%
  ggplot(aes(x = Month, y = Approval, group = 1)) +
  geom_line() +
  geom_point() +
  scale_y_continuous(
    labels = scales::percent_format(accuracy = 1),
    limits = c(0.35, 0.45)
  ) +
  theme_minimal() -> approval_plot

df %>%
  mutate(Month = factor(Month, levels = Month)) %>%
  ggplot(aes(x = Month, y = Price.per.gallon, group = 1)) +
  geom_line() +
  geom_point() +
  ylab("Price per gallon") +
  scale_y_continuous(
    labels = scales::dollar_format(),
    limits = c(3, 5)
  ) +
  theme_minimal() -> price_per_gallon_plot

grid.arrange(
  approval_plot,
  price_per_gallon_plot,
  nrow = 2,
  top = "Gas prices vs. Biden approval")
```

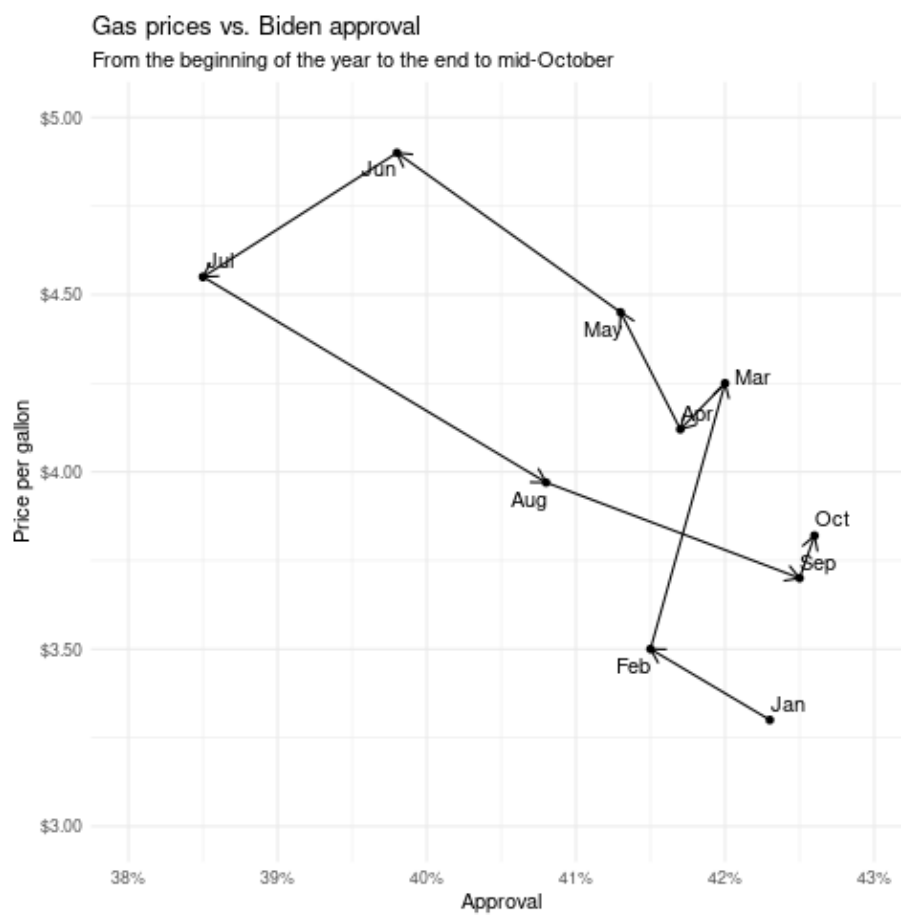
### 3.1.1 Uzasadnienie

Oryginalny wykres może być mało czytelny dla niewprawionego czytelnika. Zazwyczaj czas reprezentowany jest na osi  $x$ , na tym wykresie obie osie obrazują zmienne których zachowanie badane było w czasie.

Celem tego wykresu jest pokazanie relacji dwóch zmiennych w czasie. To samo można osiągnąć poprzez narysowanie dwóch wykresów, po jednym dla każdej zmiennej. W ten sposób bardzo łatwo jest stwierdzić czy zmiana ceny paliwa jest skorelowana ze zmianą poparcia dla prezydenta Bidena.

## 3.2 Oryginalny typ wykresu z naniesionymi poprawkami

Wykres 3 prezentuje oryginalny wykres z naniesionymi poprawkami.



Rysunek 3: Oryginalny wykres z poprawkami zwiększającymi czytelność

Kod użyty do przygotowania wizualizacji:

```
library(ggplot2)
library(ggrepel)
library(dplyr)

df <- read.csv("biden.csv")

df %>%
  ggplot(aes(x = Approval, y = Price.per.gallon, label = Month)) +
  geom_point() +
  geom_text_repel() +
  geom_segment(
    aes(
      xend = c(tail(Approval, n = -1), NA),
      yend = c(tail(Price.per.gallon, n = -1), NA)
    ),
    arrow = arrow(length = unit(0.3, "cm"))
  ) +
  labs(
    title = "Gas prices vs. Biden approval",
    subtitle = "From the beginning of the year to the end to mid-October",
    x = "Approval",
    y = "Price per gallon"
  ) +
  scale_x_continuous(
    labels = scales::percent_format(accuracy = 1),
    limits = c(0.38, 0.43)
  ) +
  scale_y_continuous(
    labels = scales::dollar_format(),
    limits = c(3, 5)
  ) +
  theme_minimal()
```

### 3.2.1 Uzasadnienie

Wykres ten jest lepszy od oryginalnego z następujących powodów. Zakres osi  $x$  został zmniejszony, dzięki czemu wykres wypełnia lepiej dostępną przestrzeń, stając się bardziej czytelny. Wprowadzony został tekstowy opis punktów, pozwalający łatwiej porównywać 2 obserwacje ze sobą. Wcześniej porównanie dwóch miesięcy oznaczało odliczenie punktów od pierwszej / ostatniej obserwacji. Zostały również dodane strzałki, które pokazują w którym kierunku znajduje się obserwacja dla kolejnego miesiąca.