

Praca domowa 2

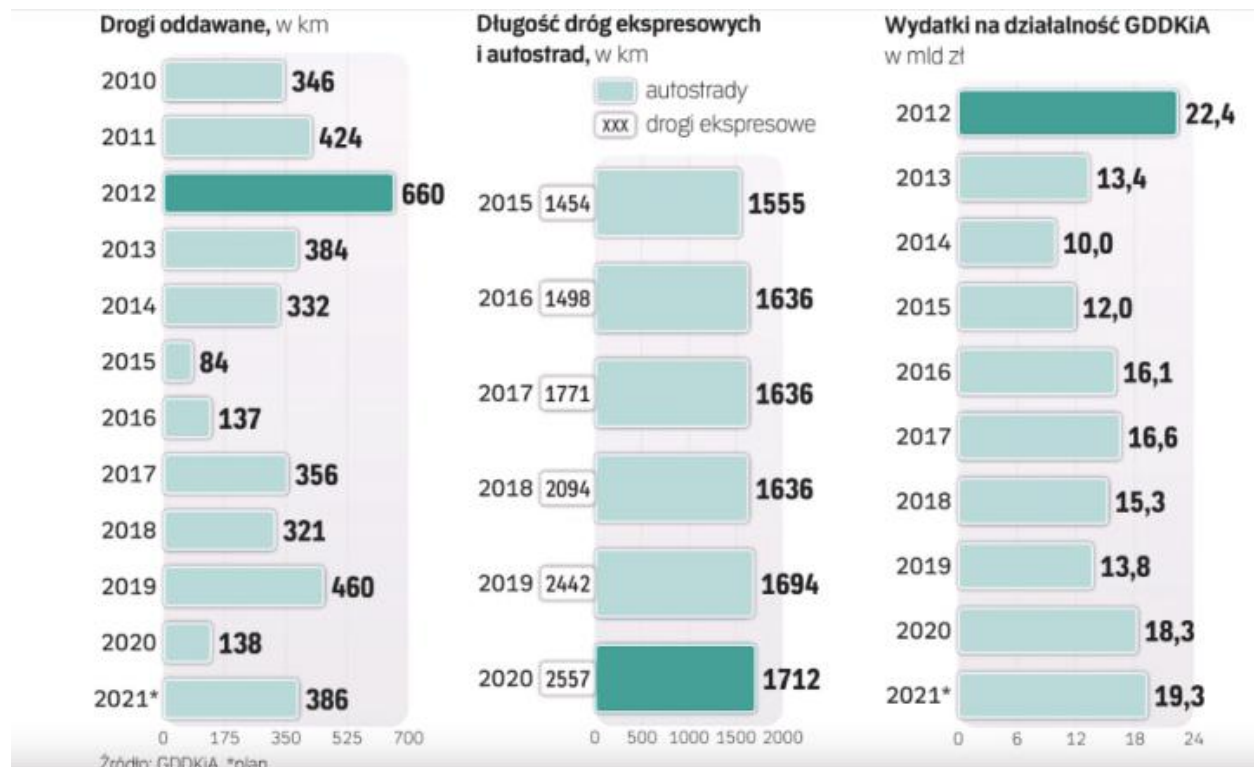
Paweł Szymański

1. Cel zadania domowego

Celem zadania domowego było znalezienie wadliwej wizualizacji danych i jej poprawienie z użyciem pakietu ggplot2.

2. Wybrana wizualizacja

Do zadania wybrano wykres ze strony: <https://www.rp.pl/Budownictwo/303219941-Mamy-zatrzymac-budowy-niezgodne-z-prawem-UE.html>.



Powyższy zestaw 3 wykresów słupkowych ma kilka znaczących problemów, które utrudniają intuicyjne zrozumienie prezentowanych informacji. Przede wszystkim wykresy te opisują różne przedziały czasu, i przedstawiane wartości dla tego samego roku nie występują koło siebie. Ponadto na środkowym wykresie użyto wykresu słupkowego z zaokrąglonymi krawędziami do porównania podobnych wartości, co utrudnia zauważenie zmiany między kolejnymi latami. Natomiast długość dróg ekspresowych nie została w żaden sposób zaprezentowana graficznie, zostały tylko wypisane wartości dla kolejnych lat. Co więcej pierwszy wykres przedstawia bardzo podobne informacje co drugi: np jeśli dodamy informacje o długości pozostałych dróg do drugiego wykresu, pierwszy wykres będzie przedstawiał tylko przyrost z wartości na drugim wykresie, bez podziału na poszczególne rodzaje dróg. Na koniec zaznaczenie na ciemniejszy kolor słupków o największej

wartości nie wnosi nic nowego, a w przypadku danych które stale rosną (wykres drugi) jest kompletnie bez sensu.

3. Propozycja poprawionej wizualizacji

Dane zebrano z oficjalnej strony GDDKiA i umieszczono je w tabeli danych.

```
rok<-2012:2021
rok_label<-paste(rok[1:10])
rok_label[10]='2021*'
autostrady<-c(296.6,125.1,76.3,20.8,81,0,0,57.7,18.3,39.6)
ekspresowe<-c(343,151,215.8,34.9,43.8,273.4,323.0,373.6,104.8,308.2)
krajowe<-c(20.4,107.9,40.4,32.3,12.4,64,17.1,28.6,14.6,37.7)
wydatki<-c(22.4,13.4,10,12,16.1,16.6,15.3,13.8,18.3,19.3)

drogi<-data.frame(rok,autostrady,ekspresowe,krajowe,wydatki)
```

Ze względu na brak wcześniejszych danych jako przedział czasu wybrałem lata 2012-2021. Wykresy całkowitej długości dróg danego rodzaju oraz rocznej oddanej długości dróg zostały zastąpione roczną długością oddanych dróg z podziałem na typy, co łączy prawie wszystkie informacje zawarte w obu tych wykresach. Wykres oddanych dróg i wydatków, znajdują się jeden pod drugim co umożliwia szybkie porównanie wartości obu w danym roku. Informacja w którym roku badane wartości były największe została usunięta, gdyż nie wnosi ona nowych informacji, od razu widoczna jest wartość maksymalna na obu wykresach.

Podsumowując poniższa wizualizacja jest czytelniejsza i stanowi ułatwienie w ewentualnej dalszej interpretacji danych.

```
library(ggplot2)

g1<-ggplot(drogi)+
  geom_bar(aes(x=rok,y=autostrady+ekspresowe+krajowe,fill='Autostrady'),stat = "identity")+
  geom_bar(aes(x=rok,y=ekspresowe+krajowe,fill='Drogi ekspresowe'),stat = "identity")+
  geom_bar(aes(x=rok,y=krajowe,fill='Drogi krajowe'),stat = "identity")+
  labs(y='Długość w km',fill='Legenda:')+
  scale_x_continuous(name='',breaks=rok, labels=rok_label, limits=c(2011.5,2021.5))+
  theme(legend.position="top")

g2 <- ggplot(drogi)+
  geom_line(aes(x=rok,y=wydatki),color='brown',lwd=2)+
  scale_x_continuous(name='Rok (* to plan na rok 2021)', breaks=rok,
                    labels=rok_label, limits=c(2011.5,2021.5))+
  ylim(0,24)+
  labs(y="Wydatki w mln")

library(grid)
library(gridExtra)

grid.arrange(g1,g2,heights=c(2,1),
             top = textGrob(paste("Długość oddanych dróg z podziałem na rodzaje\n",
                                "oraz wydatki GDDKiA we wskazanych latach",sep=""),
                           gp=gpar(fontsize=14)))
```

Długość oddanych dróg z podziałem na rodzaje oraz wydatki GDDKiA we wskazanych latach

Legenda: ■ Autostrady ■ Drogi ekspresowe ■ Drogi krajowe

