Piotr Wilkosz, GP02

1.

Dane przedstawiają informacje dotyczące kosztów ubezpieczenia w USA. Możemy z nich wyróżnić 2 typy zmiennych, którymi są:

Zmienne jakościowe: **sex**(płeć), **smoker**(informacje o paleniu), **region**(obszar zamieszkania).

Zmienne ilościowe: **age**(wiek), **bmi**(wskaźnik masy ciała), **children**(ilość dzieci objętych ubezpieczeniem zdrowotnym), **charges**(indywidualne koszty leczenia).

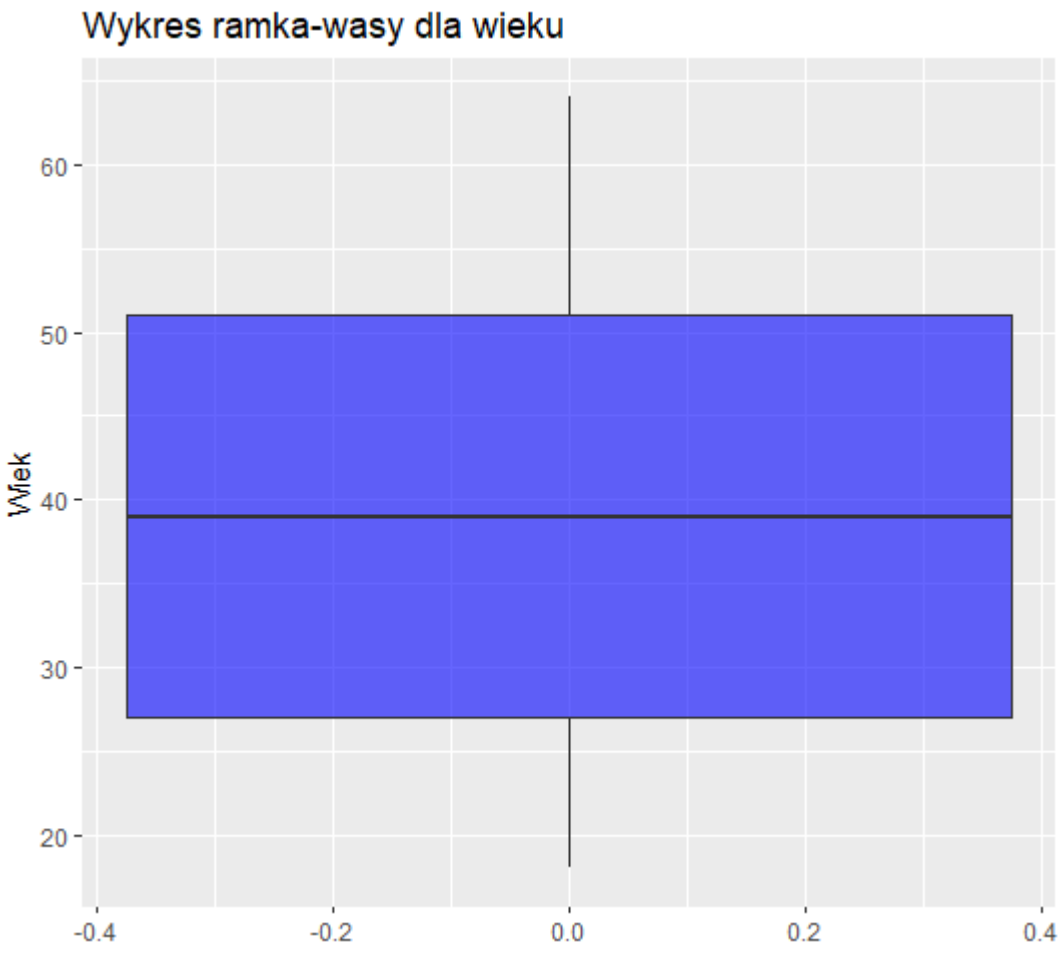
2.

Zmienna wybrana do analizy: **age**(wiek)

a) Wartości statystycznych miar położenia i rozproszenia

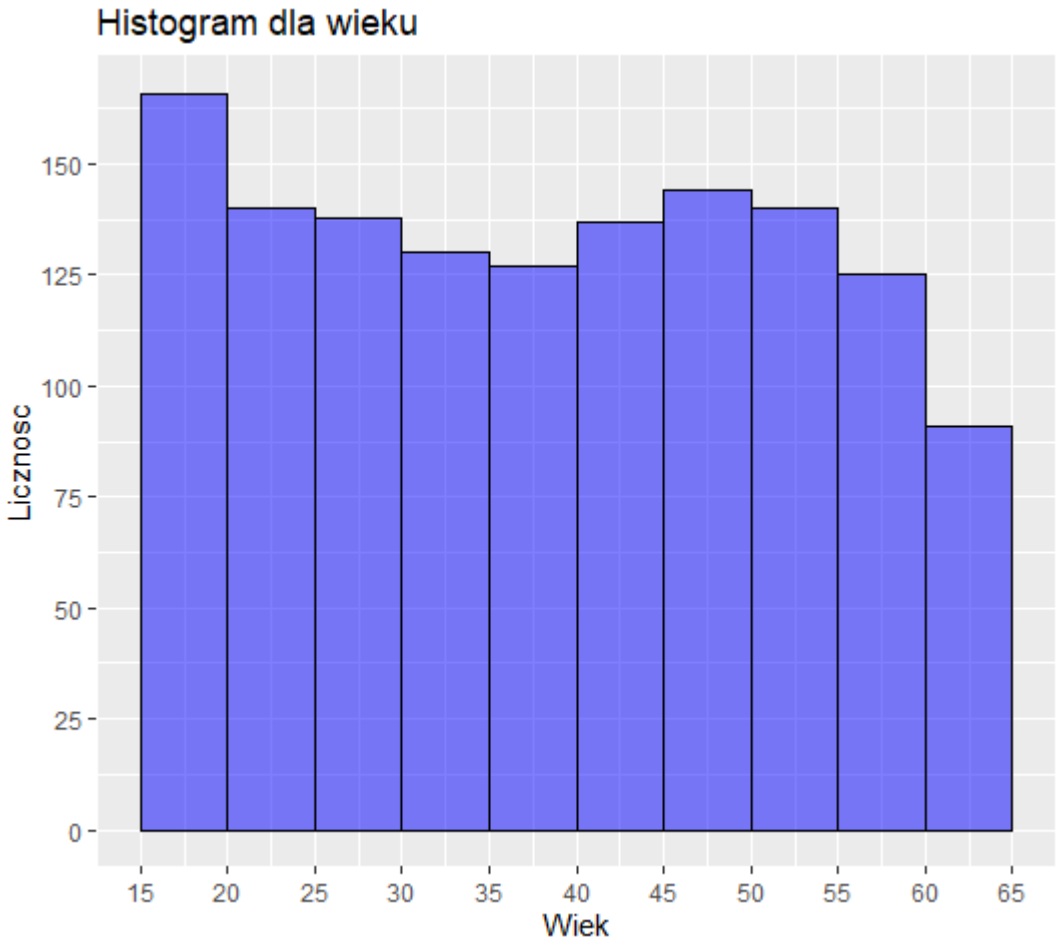
Średnia	Mediana	Moda	Dolny Kwartyl	Górny Kwartyl	Minimum	Maksimum	Rozstęp	Wariancja	Odchylenie standardowe
39.21	39	18	27	51	18	64	46	197.40	14.05

b) Wykres ramka wąsy dla mediany wieku



Dla powyższego wykresu mamy mały rozrzut danych, ponieważ na wykresie ramka-wąsy nie znajdziemy elementu który odstawał by od 1.5IQR. Mediana jest mniej więcej na środku pudełka. Dane są nierozrzucone. Oba wąsy są podobnej długości. Wykres potwierdza brak wartości odstających, potwierdza to brak kropek na wykresie.

c) Histogram dla wieku



d) Krótka interpretacja uzyskanych wyników

Na podstawie uzyskanych wyników można stwierdzić, że jest podobna ilość ankietowanych w populacji ze względu na wiek. Największa dysproporcja jest dla osób powyżej 60 roku życia. Największą grupę osób stanowią osoby w wieku 18 lat, ponieważ taka jest moda. Średni wiek ankietowanych osób wynosił 39 lat z odchyleniem standardowym 14.05 lat. Mediana wynosi 39 co oznacza, że 50% obserwacji znajduje się poniżej mediany i 50% powyżej mediany. Mediana jest zbliżona do średniej co oznacza mały rozrzut danych. Rozkład zmiennej jest w miarę symetryczny, nie widać dużej skośności.

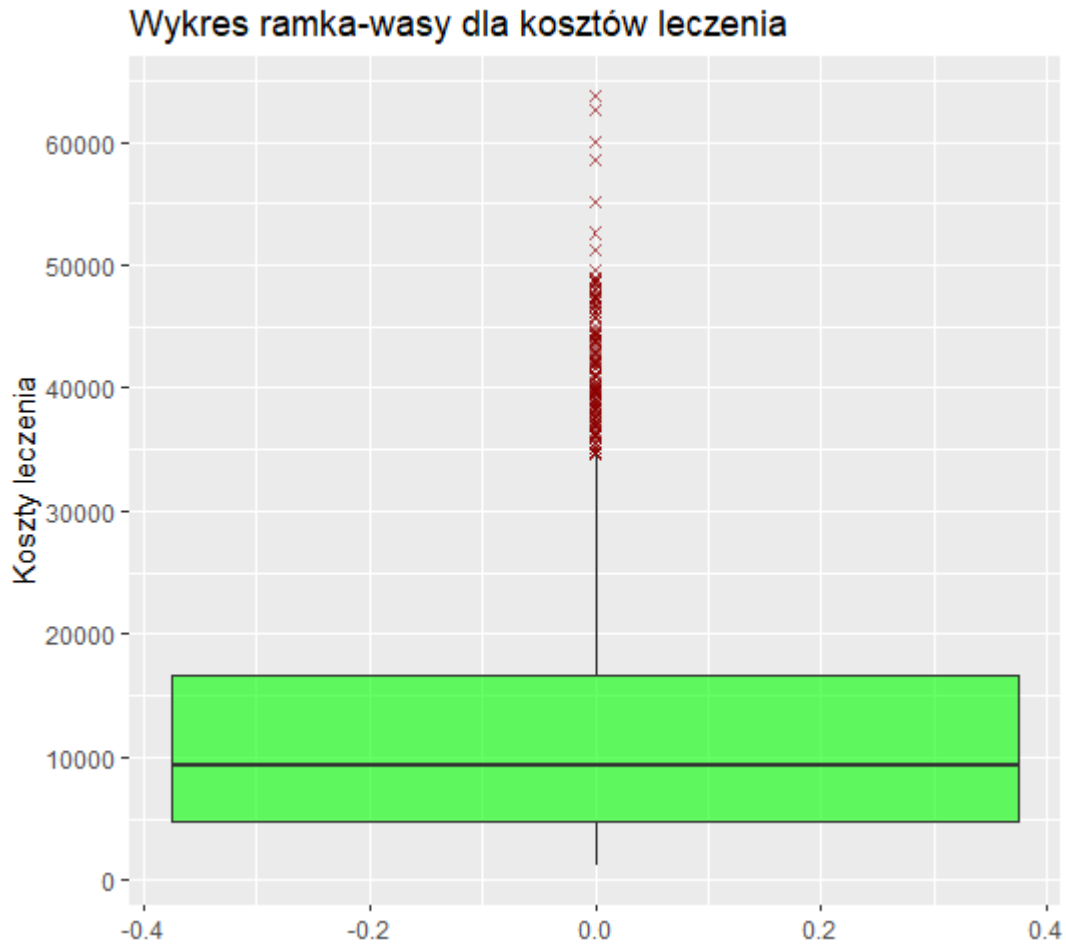
Zmienna wybrana do analizy: **charges**(indywidualne koszty leczenia rozliczane przez ubezpieczenie zdrowotne)

a) Wartości statystycznych miar położenia i rozproszenia

Średnia	Mediana	Moda	Dolny Kwartyl	Górny Kwartyl	Minimum	Maksimum	Rozstęp	Wariancja	Odchylenie standardowe
13270.42	9382.03	1639.56	4740.29	16639.91	1121.87	63770.43	62648.55	146652372.15	12110.01

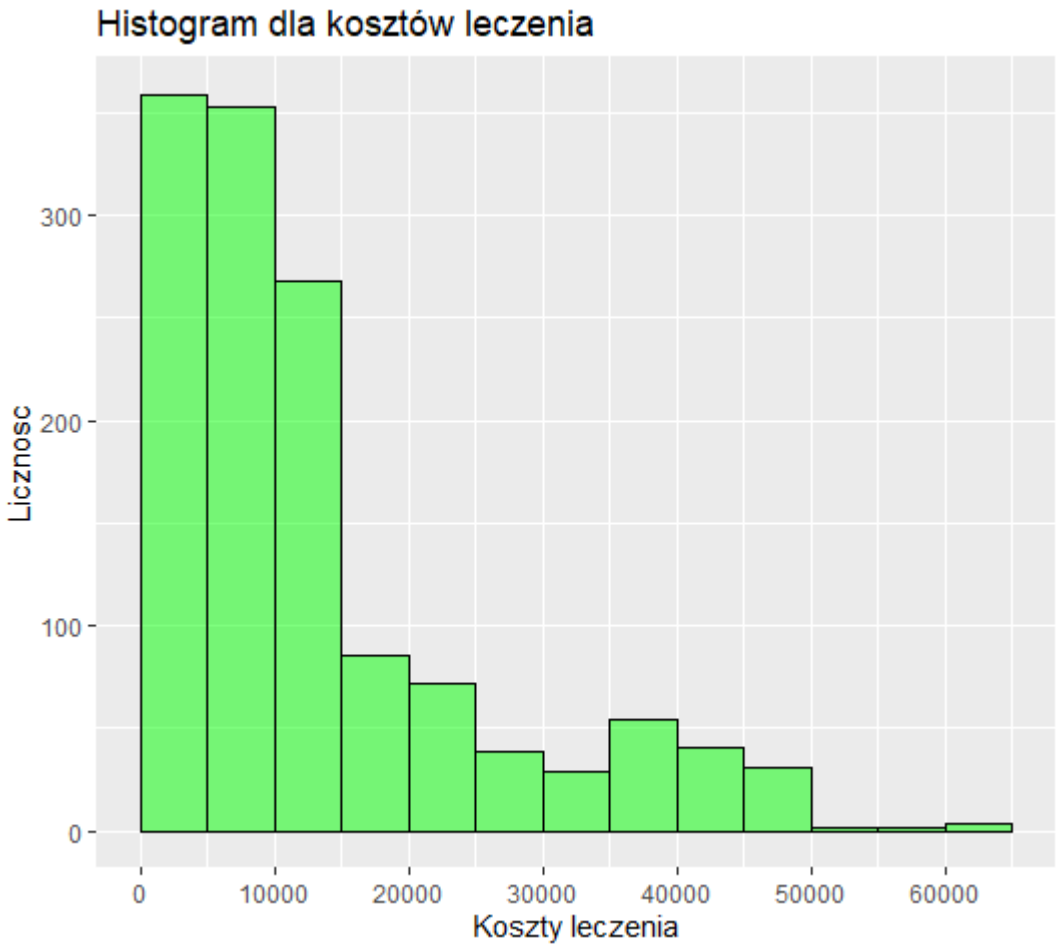
b) Wykres ramka wąsy dla mediany kosztów leczenia

c)



Dla powyższego wykresu możemy zauważyć znaczny rozrzut danych, szczególnie dla górnego kwartylu. Dolny wąs jest znacząco krótszy od górnego, co świadczy o małym rozrzucie danych poniżej dolnego kwartylu. Z tego wynika że obywatele przeznaczają na służbę zdrowia podobne kwoty, lecz istnieją zamożniejsze osoby które wydają dużo więcej co pokazują odstające wartości górnego kwartylu.

d) Histogram dla kosztów leczenia



e) Na podstawie uzyskanych wyników można stwierdzić, że najczęściej pacjenci wydają na leczenie kwoty z przedziału od 0 do 5000 dolarów. Średnia ilość wydawanych pieniędzy wynosi 13270.42 $. Najczęściej wydawaną kwotą na leczenie jest 1639.56$, ponieważ taka jest moda. Górny kwartyl przekracza 16 tyś $(skośność wykresu) co oznacza, że wydatki powyżej tej kwoty zdarzają się rzadko. Maksymalna kwota pieniędzy wydawanych na leczenie wynosiła w przybliżeniu 63770$. Minimalna kwota wydawanych pieniędzy wyniosła w przybliżeniu 1121$, co skutkuje bardzo wysokim rozrzutem wynoszącym 62648$. Wykres jest asymetryczny, prawostronnie skośny.