

1. Skompiuj poniższy kod

```
x<-c(-9,-8,-1,0,1,2,2,2,2,2,3,3,4,4,4,5,5,5,5,5,6,6,7,7,8,8,9,12,12,14,19,31)
```

```
q75=quantile(x,0.75)
```

```
q25=quantile(x,0.25)
```

```
IQR=q75-q25
```

```
plot(x,rep(1,length(x)),main="Wykres pudełkowy")
```

```
abline(v=median(x),col=2)
```

```
abline(v=min(x),col=3)
```

```
abline(v=max(x),col=3)
```

```
abline(v=q25,col=4)
```

```
abline(v=q75,col=4)
```

```
abline(v=q75+1.5*IQR,col=6)
```

```
abline(v=q25-1.5*IQR,col=6)
```

Następnie dorysuj wykres pudełkowy w wersji poziomej dla wektora x poleceniem
`boxplot(x, horizontal=TRUE, at=1, add=TRUE, axes=FALSE)`

2. Narysuj wykres pudełkowy wektora `y<-c(3, 7, 10, 4, 7, 6, 16, 1, 3, 5, 11, 11,4)`. Nanieś na nim linie wyznaczające podstawowe statystyki tej próby (min, max, mediana, 1 i 3 kwartył).

3. Zapisz lokalnie (jako ramkę danych) dane iris.

(a) Jakie dane zawiera ta baza? Które kolumny są typu factor?

(b) Korzystając z komendy `cut` utwórz nową kolumnę, która podzieli obserwacje z kolumny `Sepal.Length` na poziomy `a,b,c` itd. w punktach całkowitych, np. przedziały od `[1,2)`, `[2,3)` etc. Wybierz przedziały tak, aby skrajne z nich nie były puste.

(c) Sprawdź, jakiego typu obiektem jest kolumna utworzona w poprzednim podpunkcie. Wypisz tabelę liczości dla tej kolumny. W którym przedziale jest najwięcej obserwacji?

(d) Wypisz tabelę krzyżową względem dwóch zmiennych typu factor z tej ramki.

4. Dla danych `state.x77` (zapisanych lokalnie jako ramka danych stany) narysuj histogram dochodów na osobę (`Income`). Ustaw zakres na osi OY na 0-20 i dorysuj do wykresu poziomy wykres pudełkowy dla tych dochodów.

(a) Czy mediana dochodów znajduje się w najliczniejszym przedziale histogramu?

(b) Czy są jakieś obserwacje odstające? Jeśli tak, to które to stany?

5. Skorzystaj z funkcji `cut`, aby do danych stany z poprzedniego zadania dodać nową kolumnę `IlliteracyLevel`, która przydzieli poszczególnym stanom poziom analfabetyzmu: `low` dla przedziału `(0,1]`, `medium` dla `(1,2]` i `high` dla wartości powyżej 2. Narysuj wykres `plot()` dla tej zmiennej. Ile jest stanów w kategorii `high`? Wypisz ich nazwy.

6. Ustawienie argumentu `notch=T` w wykresie pudełkowym powoduje narysowanie na pudełku wcięcia, obrazującego 95% przedział ufności dla mediany (czyli z 95-procentową pewnością mediana będzie gdzieś w tym przedziale). Jeśli wcięcia dwóch lub więcej wykresów pudełkowych nie nakładają się na siebie (w poziomie), sugeruje to, że mediany się różnią. Dla bazy stany z poprzedniego zadania narysuj wykres pudełkowy z wcięciami rozkładu zmiennej `stany$Income` w zależności od `stany$IlliteracyLevel` i na ich podstawie oceń, czy mediany dla poszczególnych grup są równe. Skomentuj, czy poziom analfabetyzmu wpływa na wysokość dochodu na osobę.