J2P3BM & J2P3BT Oefententamen Data-analyse in R





Oefententamen J2P3 Data-analyse in R

Aan het einde van het tentamen lever je een script in waarin alle opdrachten uitgewerkt staan. Zorg voor een duidelijk script, waarin de verschillende opdrachten zijn weergeven achter een #. Hieronder zet je de code van de uitvoering (zoals in onderstaande voorbeeld is weergeven). Voor dit oefententamen hoef je niks in te leveren, maar oefen wel alvast met het maken van een script voor het tentamen.

```
# Opdracht 1
appel <- c(1, 1)
peer <- c(2, 3)

# Opdracht 2
appel_peer <- appel + peer
# Uitkomst bekijken
appel_peer</pre>
```

De dataset die je voor dit oefententamen gebruikt bestaat uit gegevens over ontbijtgranen. De dataset is op BlackBoard te vinden onder de naam *cereal.csv*. In deze dataset staan onder andere informatie van welk merk bepaalde ontbijtgranen zijn, wat voor (en hoe veel) voedingsstoffen erin zitten, en hoe lekker ze gevonden worden.

Opdracht 1 (2 punten)

Bekijk de scheidingstekens van *cereal.csv* met behulp van een tekst editor, en of *headers* aanwezig zijn. Laad de data vervolgens in R onder een zelfgekozen naam (1 punt). Test of het inladen goed is gegaan door de bovenste 10 rijen van de data te bekijken (1 punt).

Opdracht 2 (2 punten)

Download zowel *ggplot2*, als *dplyr* (1 punt). Laad ze hierna ook in R zodat je de functies uit deze packages kan gebruiken (1 punt).

Opdracht 3 (4 punten)

Bekijk welke ontbijtgranen warm gegeten horen te worden. Bekijk dit door een subset/selectie van de data te maken (1 punt) op basis van de kolom type, waarin **C** staat voor ontbijtgranen die koud gegeten horen te worden en **H** voor ontbijtgranen die warm gegeten horen te worden (1 punt). Sla de uitkomst op onder een nieuwe, zelfgekozen naam.

Sorteer de uitkomst van alleen de ontbijtgranen die warm gegeten worden op basis van de kolom met scores (rating) (1 punt). Gebruik hiervoor order() (1 punt).

Opdracht 4 (5 punten)

Maak een pipe (%>%), waarin je alle data selecteert welke bij het merk *Kelloggs* horen. Hiervoor filter je op **K** (voor Kelloggs) in de kolom mfr. Sla de gefilterde tabel op onder een nieuwe, zelfgekozen naam. (1 punt)

Maak een histogram met behulp van ggplot2, waarbij je de scores (kolom rating) van Kelloggs ontbijtgranen op de x-as zet en als opvulkleur de plankhoogte (kolom shelf) weergeeft (1 punt). Je gebruikt hiervoor dus de tabel die je eerder in deze opgave hebt gemaakt om te kijken of hogere scores vaker op andere plankhoogtes staan. Let hier wel bij op dat je de plankhoogte als factor aan moet geven, omdat deze als getallen weergeven zijn (1 punt). Geef bij de histogram aan dat je een binwidth van 20 wil gebruiken (1 punt). en dat de staven van de plankhoogtes naast elkaar komen te staan (1punt).

Opdracht 5 (4 punten)

Het kan interessant zijn om de hoeveelheid eiwit in de verschillende ontbijtgranen te analyseren. Gebruik hiervoor een pipe. Niet alle informatie van de verschillende ontbijtgranen staan per hetzelfde gewicht weergeven. Gebruik mutate() (1 punt) binnen je pipe om een nieuwe kolom protweight toe te voegen waarin je de hoeveelheid eiwit (kolom protein) deelt door het gewicht waarin alle informatie is weergeven (kolom weight) (1 punt).

Analyseer protweight met behulp van summarize() binnen je pipe (1 punt). Zorg ervoor dat je het minimum, maximum en het gemiddelde weergeeft in je output van de pipe (1 punt).

Opdracht 6 (5 punten)

Maak een puntdiagram voor de gehele dataset, waarin je de hoeveelheid suiker op de x-as zet, het aantal calorieën op de y-as (1 punt) en de rating gebruikt voor de grootte van de puntjes (1 punt). Geef in de puntdiagram aan dat de punten wat doorzichtig moeten zijn (een alpha van 0.4), en de kleur darkolivegreen4 moeten krijgen (1 punt). Gebruik facet_grid om de merken apart te plotten in rijen (1 punt). Sla de plot op onder een zelfgekozen naam en bekijk de plot die je gemaakt hebt (1 punt).

Opdracht 7 (4 punten)

Als laatste ga je de plot die je bij opdracht 6 hebt gemaakt nog verder opmaken. Voeg hiervoor het onderstaande toe:

- Een zelfgekozen titel (1 punt)
- Verander de labels van de legenda naar een zelfgekozen naam (1 punt)
- De al bestaande theme_gray (1 punt)
- Een eigen toevoeging aan de opmaak (met theme), waarin je de getallen op de assen (axis.text) verandert naar de kleur darkolivegreen4. (1 punt)

