

R-seminar 4: Oppgaver

STV1020 Vår 2021

Uke 14

1. Last inn datasettet "VALGDATA.Rdata" og last inn pakken tidyverse.
2. Hva er navnene på variablene i datasettet?
3. Hvor mange missing er det totalt i datasettet?
4. Hvor mange missing er det på hver enkelt variabel?
5. Lag et subset av datasettet hvor du fjerner NA på variabelen valg.
6. Hvor mange kvinner og hvor mange menn er det i datasettet?
7. Gjør variabelen alder til en numerisk variabel og kjønn til en factor variabel.
8. Finn gjennomsnittsalderen til menn og deretter kvinner. Regn ut forskjellen.
9. Få oversikt over variabelen valg. Lag et plot som viser den univariate fordelingen. Hvor mange har stemt på hvert av partiene?
10. Finn korrelasjonen mellom alder og rik.
11. Lag et spredningsdiagram mellom alder og rik med lineær støttelinje. Endre navnene på x-aksen, y-aksen og gi diagrammet en tittel. Tolk form, retning og styrke.