

Inlämningsuppgift 1

Instruktioner

1. Lös uppgifterna 1-3 på följande sidor *i grupp* och lämna in följande i Canvas:
 - Ett Quarto-dokument med filnamn (ex. för Grupp1): Grupp1.qmd där ni svarar på frågorna i Uppgift 1-3 och skapar diagrammet i Uppgift 3. **Kom ihåg att ange namn på dem som deltagit i arbetet.**
 - Ett pdf-dokument som är kompilerat från qmd-filen (eller sparat som pdf från en html- eller word-version av Quartodokumentet).
2. Gör *individuellt* kapitlet “Introduction” i kursen “Introduction to Data Visualization with ggplot2” i DataCamp.

Uppgift 1

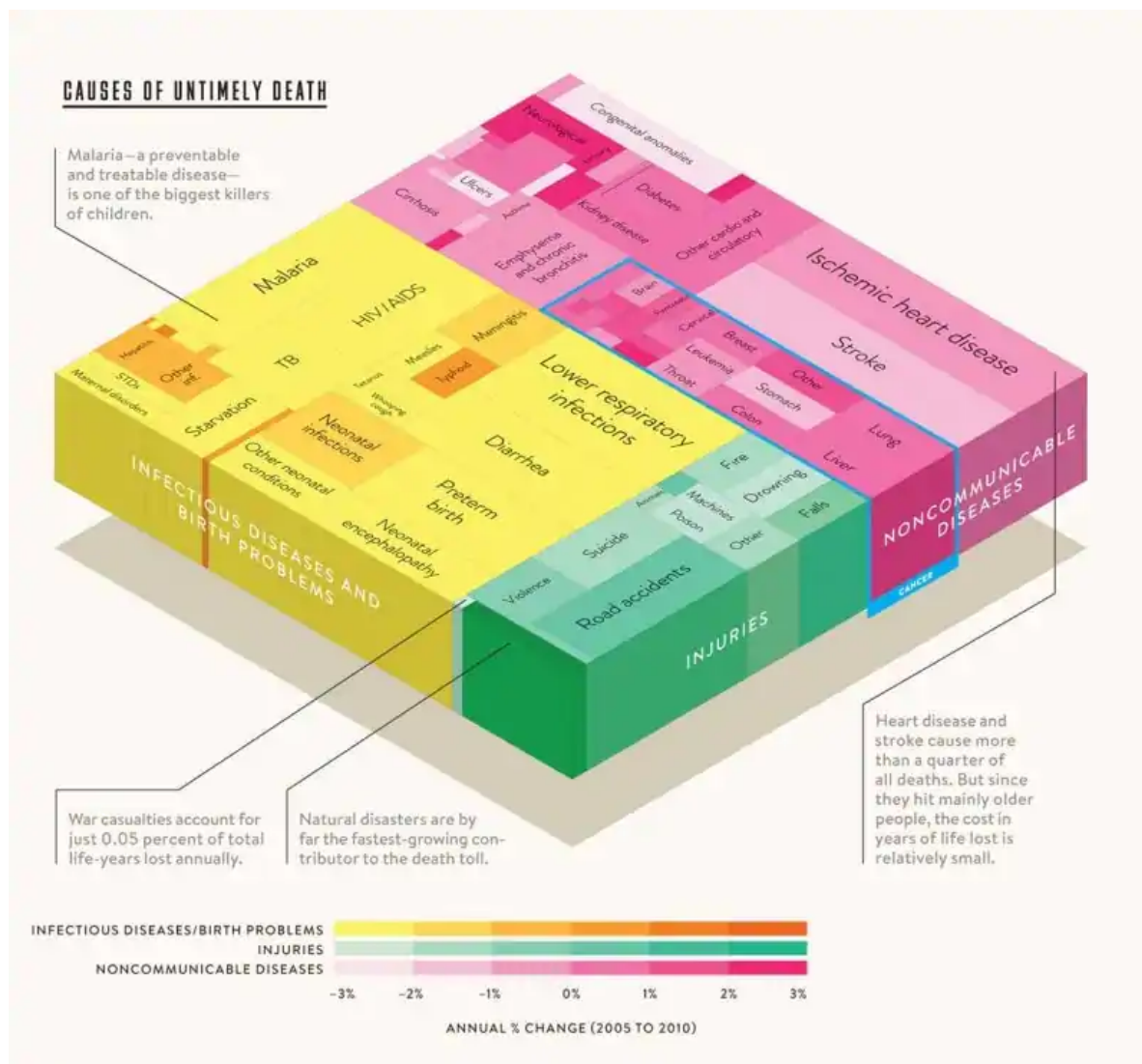
I Washington Post (2013-12-27) finns en artikel där Microsofts grundare Bill Gates presenterar sitt favoritdiagram från året (Figur 1).

Bill Gates säger följande: "I love this graph because it shows that while the number of people dying from communicable diseases is still far too high, those numbers continue to come down. In fact, fewer kids are dying, more kids are going to school and more diseases are on their way to being eliminated. But there remains much to do to cut down the deaths in that yellow block even more dramatically. We have the solutions. But we need to keep the up support where they're being deployed, and pressure to get them into places where they're desperately needed."

Se t.ex. <https://en.wikipedia.org/wiki/Treemapping> för en beskrivning av vad en treemap är.

Frågor att besvara:

- a) Vilka frågor besvarar visualiseringen?
- b) Vilka variabler beskrives med hjälp av
 - storlek på areor
 - färger
 - färgtoner (mörka och ljusa toner)?
- c) Hur påverkar 3D-effekten möjligheten att jämföra areornas storlekar?
- d) Beskriv ytterligare styrkor/svagheter med visualiseringen utifrån olika principer om visualisering.
- e) Kan du föreslå några förbättringar? Hur skulle detta kunna ha gjorts bättre?



Figur 1: Treemap: dödsorsaker

Uppgift 2

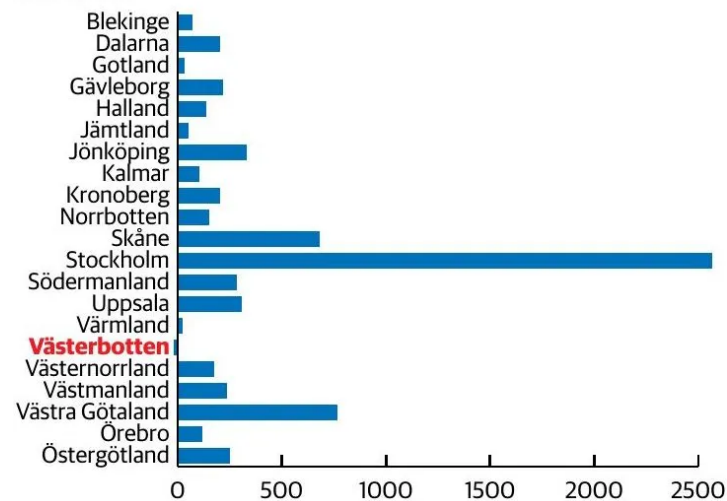
I Västerbottens-Kuriren (VK) rapporterades den 16 januari 2021 att Västerbotten som enda region hade en underdödlighet under 2020 (artikel bifogad separat). Läs igenom artikeln och besvara sedan frågorna nedan.

Frågor att besvara:

- Hur definieras över-/underdödlighet för år 2020 enligt Statistiska centralbyrån (SCB)?
- I Figur 2 (från VK-artikeln) nedan kan man se att Stockholm, Västra Götaland och Skåne var de tre regioner med högst överdödlighet under perioden 2 mars - 27 december 2020. Vad tror du att det kan bero på?
- Föreslå ett alternativt sätt att skapa ett stapeldiagram som på ett mer "rättvist" sätt jämför överdödligheten i Sveriges regioner 2020.

Överdödlighet under pandemin

Så många har dött i Sveriges län under perioden 2 mars-27 december 2020 jämfört med genomsnittet under motsvarande period åren 2015-2019.



Figur 2: Överdödlighet per region

Uppgift 3

På SCB:s hemsida kan man hitta statistik om genomsnittlig månadslön efter kön och utbildningsnivå för år 2021. För statistik om löner i Sverige, se nedanstående länk.

<https://www.scb.se/hitta-statistik/sverige-i-siffror/utbildning-jobb-och-pengar/>

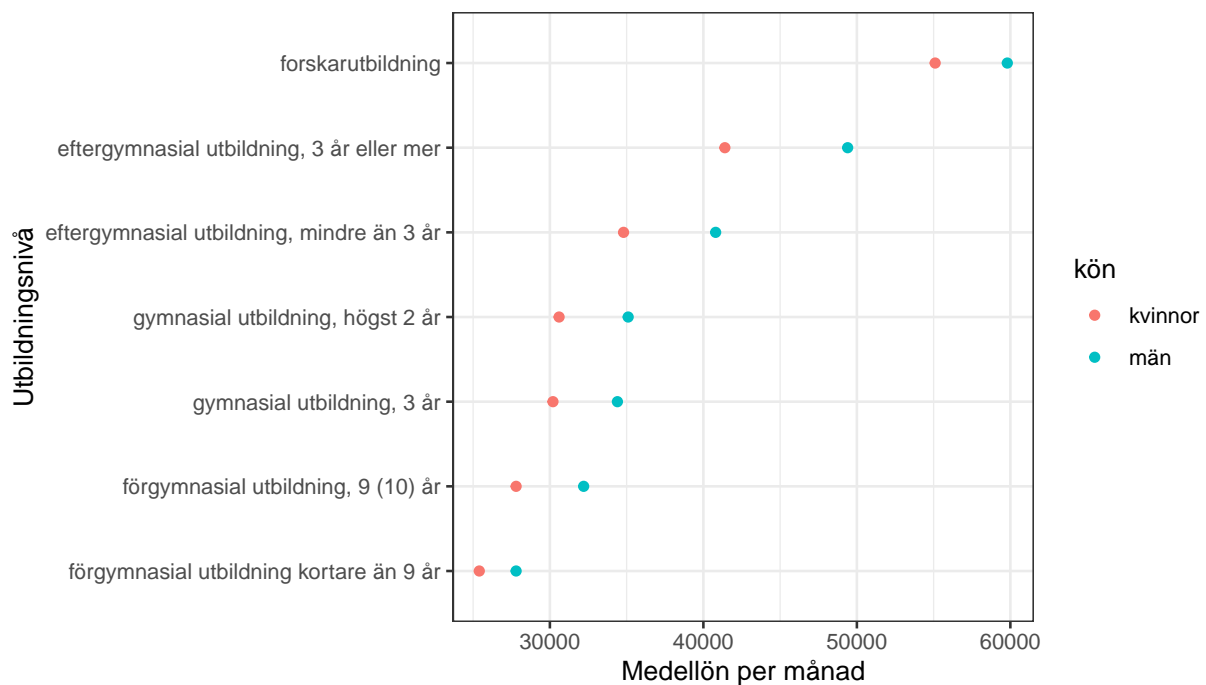
Datamaterialet `MedellönUtbildningKön2021.xlsx` finns bifogat till inlämningsuppgiften.

Skapa ett punktdiagram

- Läs in data i R (använd t.ex. paketet `readxl`)
- Gör om data till 'long format' med hjälp av `tidyr`
- Återskapa Figur 3 med hjälp av `ggplot2`

Frågor att besvara:

- Följer medellönen utbildningsnivå eller finns undantag?
- Hur påverkas mottagaren av att skalan inte börjar på 0?
- Vilka ändringar eller tillägg skulle du göra om detta diagram skulle publiceras i en dagstidning?



Figur 3: Medellöner