Data inlezen en ggplot2 - Huistaak

Herhaling lesmateriaal

- Download het materiaal van de tweede les
- Doorloop het document rond **Data inlezen**, eventueel met behulp van het script
  - Noteer alle onduidelijkheden en mail die naar Pieter (pieter.verschelde@inbo.be) ern Raïsa (raisa.carmen@inbo.be)
- Doorloop het document rond **ggplot2**, eventueel met behulp van het script
  - Indien er iets niet duidelijk is, bekijk de help van de desbetreffende `geom_xxx()`.
  - Geraak je er nog niet wijs uit, mail dan je vraag met het stukje code naar Pieter en Raïsa

Lees een eigen dataset in

- Probeer enkele van je eigen datasets in te lezen. Doe dit vanuit het project dat je in de huistaak van R en RStudio aangemaakt hebt. Kies minstens 2 formaten uit
  - csv
  - txt
  - Excel
  - Googlesheet

Maak je vertrouwd met de `pilootstudie` data

Pilootstudie

- 8 proefvlakken (4 eik en 4 beuk)
- 12 bomen per proefvlak
- 7 ploegen
- elke ploeg meet helft van de bomen 1x en andere helft 3x
  - wisselt per ploeg
  - indien inconsistentie tussen 3 metingen, dan wordt de boom een 4e keer gemeten
- omtrek meting
  - gemeten op borsthoogte (+/- 130cm)
  - gemeten tot op 1cm nauwkeurig
- hoogte meting
  - 2 toestellen (vertex, fieldmap)
- referentie omtrek
  - gemiddelde van de 3 metingen van ploeg 7
  - gemeten op exact 130cm hoogte
  - gemeten tot op 1mm nauwkeurig

1. Lees de gegevens in van het bestand *pilootstudie.csv* in de data folder.

2. Bekijk de structuur en controleer of alle variabelen van het correcte datatype zijn.
3. Vraag een `summary()` van de gegevens.

    a. Klopt het aantal proefvlakken?
    b. Klopt het aantal ploegen?
    c. Hoe zijn de bomen genummerd?
    d. Wat is het maximum aantal metingen per boom?
    e. Zijn er ontbrekende waarden voor bepaalde variabelen?
    f. In welke range liggen de omtrek en de hoogte van de bomen?

Verkennende plotjes van de `pilootstudie` data

Indien je liever vergelijkbare plotjes maakt van je eigen gegevens (die je reeds hebt kunnen inlezen), dan wordt dit zeker aangemoedigd.

1. Maak een histogram van de omtrek.
2. Verschilt de gemeten hoogte tussen de toestellen?

    a. Maak hiervoor een boxplot.
    b. Splits deze op volgens proefvlak.
    c. Kleur volgens toestel (bestudeer het verschil tussen `color` en `fill`).

3. Is er een verband tussen hoogte en omtrek?

    a. Maak hiervoor een scatterplot.
    b. Voeg een (al dan niet lineaire) smoother toe.
    c. Verander de titel in “Verband tussen hoogte en omtrek”.
    d. Verander de naam van de X-as in “Omtrek op borsthoogte (in cm)” en die van de Y-as in “Boomhoogte (in m)”.

4. Is het verband tussen hoogte en omtrek afhankelijk van het gebruikte toestel voor de hoogte?

    a. Voeg aan de vorige scatterplot een kleur toe volgens het toestel.
    b. Bereken ook een afzonderlijke smoother voor beide toestellen.

5. Is het verband tussen hoogte en omtrek afhankelijk van het proefvlak?

    a. Maak hiervoor per proefvlak een scatterplot met een rode smoother.
    b. Bewaar deze figuur in de map “Figuren/” onder de naam “OmtrekHoogte_PerProefvlak.png”.
    c. maak een scatterplot per toestel en per plot via facet_grid