

Lihtsamad tidyverse käsud (abiks kursusel HVEE.01.006)

Töömudel on järgmine

```
# andmed %>%  
# protsess1() %>%  
# protsess2()
```

Kõigepealt tuleb sisse lugeda vajalikud paketid/raamatukogud

```
# library(tidyverse)  
# library(tidytext)  
# ja veel mõni kui vaja
```

%>% - vii andmed järgmisesse protsessi

```
# select() vali muutujad  
# filter() vii andmestik läbi filtri/sõela  
# arrange() järjestada andmed  
# group_by() grupeeri andmestik  
# summarise() võtad andmed kokku mõne funktsiooniga  
# join() ühenda kaks andmestikku  
# mutate() loo uus muutuja
```

Käsud tekstitöötluseks

```
# count(muutuja) - loeb esinemiste arvu  
# top_n(number, muutuja) - võta esimesed 10 muutuja kohta  
# left_join(andmestik) - ühenda üks andmestik teisega  
# inner_join(andmestik) - ühenda üks andmestik teisega, jättes alles ainult sama väärtusega read  
# anti_join(andmestik) - eemalda esimesest andmestikust read, millel on teises tabelis samas tulbas samad read  
# unnest_tokens(ühik, muutuja) - segmenteeri tekst sõnedeks, 2-grammideks, 3-grammideks  
# unnest(muutuja) - võta andmestik pesast välja  
# rename(uusnimi=vananimi) - muuda tulba nime  
# spread(tulp laiiali laotamiseks, tulp väärtustega) - tee ühest tulbast mitu  
# Vaata ka data-transformation-cheatsheet.pdf, samuti trükkides nt ?select
```

Töö regulaaravaldistega

```
# str_detect(muutuja, "sõne") - osaline ühestamine  
# str_extract(koht, otsing) - võta leitav tekst välja  
# str_replace(koht, otsing, asendus) - asenda sõnaosa millegagi  
# str_remove(koht, otsing) - asenda otsitud sõna millegagi (str_extracti vastand)  
# Vaata ka strings.pdf abimaterjalide hulgas, samuti trükkides nt ?str_detect
```

Visualiseerimiseks saab kasutada ggplot-i

```
# Selle põhikuju tidyverse-is:  
#  
# andmed %>%  
# ggplot(mapping = aes(<MAPPINGS>)) +  
#   <GEOM_FUNCTION>()  
#  
# Vaata ?ggplot2 üksikasjalikumaks infoks. Google ggplot2 aitab ka.  
# Samuti vaata fail 8_ggplot2.R kursuse materjalides  
# Ja ka ggplot2-cheatsheet-2.1.pdf abifailide kataloogis.
```