

## Hvad er WordCounter?

WordCounter analyserer tekst og giver dig de mest benyttede ord og fraser. På den måde kan du kvantitativt opsummere din tekst, og undersøge hvordan ordene bliver brugt. Gennem korte datahistorier udvikler deltagerne deres datakompetence ved at finde og opridsse historier om hvilke ord, der dukker op i kendte musikeres tekstuniverser.

## Læringsmål

- Styrke evnen til at analysere og reflektere om tekst data.
- Forståelse for, at ét datasæt kan afføde mange historier.
- Evne at kommunikere datadrevne historier til andre.

## Øvelsen

### Problemløsning

Tekst er også data! Det er svært at finde mønstre i store datasæt uden brug af computere. WordCounter hjælper dig i gang med at analysere tekst kvantitativt ved at vise dig de mest benyttede ord og fraser.

### Hav de gode eksempler med

Visualisering af sangtekster er altid populært. Vis eksempelvis Tahir Hemphills "Rap Research Lab" (<http://rapresearchlab.com>) eller Javier Arce's "Spotimap" (<http://javier.is/spotimap>).

Spørg deltagerne hvilke historier, de ser i eksemplerne.

### Introducer værktøjet

Åben WordCounter (<https://databasic.io/wordcounter>) og vælg "xxx" fra menuen med eksempler. Tryk på "Count!" for at analysere teksterne fra "xxx". Vis deltagerne, hvordan word cloud'en fordeler sig, og gør opmærksom på, at ordet "yyy" bruges meget. Men det er kun ved at se nærmere på bigram and trigram, at du kan forstå hvordan og hvorfor. Det at "xxx" så ofte siger "yyy", men kun siger "zzz" en enkelt gang, kan være en sjov og afslørende historie at skitsere.

Nævn muligheden for at downloade CSV-filerne for yderligere analyse.

## Tidsramme

30 til 45 minutter

## Deltagere

3 -100 personer. Alder: 12+.  
*Udviklet for 6.-9. klasse, ungdomsuddannelser, nyhedsorganisationer, non-profit og frivillige organisationer. Tidligere erfaring med data er ikke nødvendig.*

## Lokale

- Projektor og computer
- Mulighed for at kunne arbejde i grupper af tre omkring en computer.
- Store borde, gulv- eller vægp-lads til placering af post-its og tegninger.

## Udstyr

- En computer pr. gruppe
- Stort papir/flip-overs
- Skriveredskaber: Store tuscher, farveblyanter o.lign.

# Øvelsen ( fortsat )

## Igangsæt øvelsen

1. Deltagerne har 15 minutter.
2. Deltagerne arbejder i 3-mandsgrupper.
3. Hver gruppe bruger WordCounter til at analysere tekster fra en musiker og finder frem til en historie om teksterne. <https://databasic.io/wordcounter>
4. Hver gruppe skitserer deres historie visuelt på et stort stykke papir vha. farveblyanter.

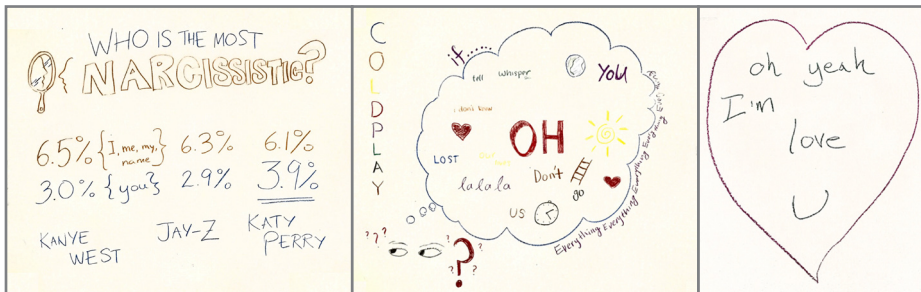
## Præsentationsrunde

Mens du går rundt blandt grupperne, fokuserer du på de forskellige typer historier, der kan udledes af datasættet. Hvis du bemærker noget særligt interessant (tegninger/visualiseringer eller enkle nedslag) husker du på dem, så du kan tage dem op under præsentationsrunden ved afslutningen af workshoppen. Med den begrænsede tid for øje spørger du efter grupper, som vil dele deres historie. Begræns hver gruppes fremlægning til 1 minut, så I ikke fortaber jer, og når hele vejen rundt. Mens grupperne deler deres viden, sørger du for, at de grupper, du tidligere bemærkede, også får tid:

Fokuser på disse emner:

- Der kan fortælles mange historier ud fra samme data
- Skitser kan være et godt hjælpemiddel, når du skal kommunikere dine idéer til folk omkring dig
- At arbejde med data kan være sjovt
- Demonstrerer hvordan en CSV-fil kan downloades og analyseres yderligere i fx Excel.

## Sample Sketches



## Husk

- Tændstik-figurer er helt fint.
- Datahistorier kan være så enkle som: "Elvis bruger ordet 'love' mere end ordet 'clambake'."
- Ved at markere boksen "stop-words", kan du enten udelade eller inkludere de mest benyttede ord som eksempelvis "en" eller "og" m.m.
- Det handler ikke om perfektion - bare skitsér noget :)

## Vigtige termer

**Bigrams** - To ord.

**csv** - En "komma-separeret-værdi"-fil ("comma-separated-values"); en standardiseret måde hvorpå tabeller eller information kan gemmes i en tekstfil.

**Kvalitative data** - Beskrivende data, fx et billede eller et citat

**Kvantitative data** - Data der tæller eller måler noget, fx et tal.

**Stopord** - Ord der med fordel kan udelades ved tekstanalyse, fx "dette", "denne", "og", "eller" osv.

**Trigram** - Tre ord.

**Word Cloud**

Visualisering af ords forekomst i en tekst. Oftest benyttede ord er store og mindst benyttede ord er små.