

● Regular expressions

Kresten, Katrine og Laurits

1. What regular expressions do you use to extract all the dates in this blurb:

<http://bit.ly/regexexercise2> and to put them into the following format

YYYY-MM-DD ?

<https://regex101.com/r/3TcMaC/1>

2. Write a regular expression to convert the stopwordslist (list of most frequent Danish words) from Voyant in <http://bit.ly/regexexercise3> into a neat stopwords list for R (which comprises "words" separated by commas, such as <http://bit.ly/regexexercise4>). Then take the stopwordslist from R <http://bit.ly/regexexercise4> and convert it into a Voyant list (words on separate line without interpunction)

Fra Voyant til R

<https://regex101.com/r/d4VptT/1>

Fra R til Voyant

<https://regex101.com/r/DGHMGZ/1>


3. Does OpenRefine alter the raw data during sorting and filtering?

OpenRefine giver forskellige måder at sortere og filtrere data på, uden at det påvirker den indsatte data. Man kan fx bruge "include" eller "exclude" funktionerne der ikke ændrer på noget af dataen, men udvælger noget af det data man vil fokusere på, og dermed gøre det mere overskueligt.

4. Fix the interviews dataset in OpenRefine enough to answer this question:

"Which two months are reported as the most water-deprived/dryest by the interviewed farmer households?"

Oktober (74) og September (70) er de måneder der er tørret ud fra dataen.



The screenshot shows a facet titled 'months_no_water' with a 'change' button. It displays 11 choices, sorted by name count. The data is as follows:

| Month | Count |
|-------|-------|
| Oct | 74 |
| Sept | 70 |
| Nov | 51 |
| NULL | 45 |
| Aug | 33 |
| Dec | 11 |
| Jan | 2 |
| July | 2 |
| Apr | 1 |
| June | 1 |
| May | 1 |

☆ Reuse This project `grel: value.replace("[", "").replace("]", "").replace(" ", "").replace("","")`

☆ Reuse This project `grel: value.split(";")`

Vi har brugt det her. Ved ikke lige, hvad vi skal kalde den:)

- 5. Real-Data Challenge: What are the 10 most frequent occupations "erhverv" among unmarried men or women of 20-30 years in 1801 Aarhus census dataset? (hint: first select either men or women to shrink the dataset to a manageable size, then filter by age, and then use merging tSo cut the erhvervvariation ruthlessly.)**

Ugifte mænd i alderen 20-30 år



Overstående er de 10 mest nævnte erhverv for ugifte mænd i alderen 20-30.

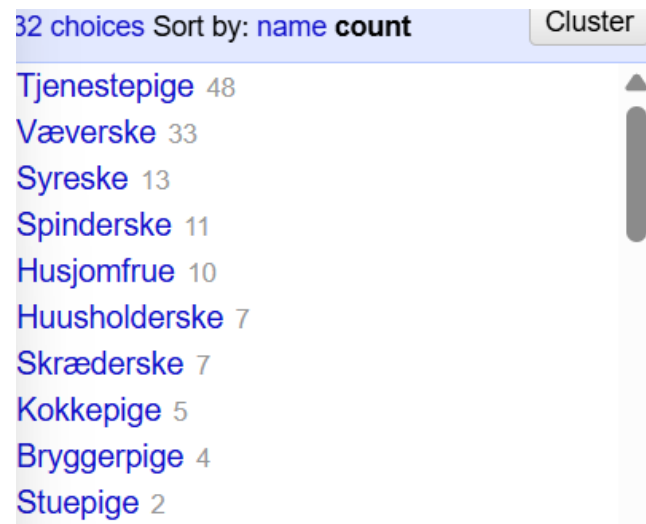
Jeg har merged/clustered meget, da der var mange variationer, men listen viser tydeligt at det største erhverv for gruppen har været en "Del af hæren" i forskellige grader.

Jeg har merged erhvervene i store træk. Jeg kunne have fortsat. Men listen viser stadig, at der er flest unge ugifte mænd, som er i hæren. → Under "En del af hæren" er mange forskellige grader. F.eks. "Soldat" og "soldat i jyske regiment" m.m.

Jeg har samtidig samlet alle "svende" i én enkelt gruppe, og det samme har jeg gjort med "væver", "tjener", "karl", "vanfør"(handicappet) og "skoleansat". På den måde har jeg til en vis grad grupperet erhvervene.

Ugifte kvinder i alderen 20-30 år

Der er i alt 2244 kvinder der passer til disse kriterier



| | count |
|---------------|-------|
| Tjenestepige | 48 |
| Væverske | 33 |
| Syreske | 13 |
| Spinderske | 11 |
| Husjomfrue | 10 |
| Huusholderske | 7 |
| Skræderske | 7 |
| Kokkepige | 5 |
| Bryggerpige | 4 |
| Stuepige | 2 |

Dette er de ti mest nævnte erhverv, der er også en del der under erhverv er blank, men kigger man under *famstand* står der en noget andet, det er ikke noget der ændre på hvad de ti mest almindelige jobs er men det ændre markant antallet af de forskellige erhverv.

Lægger man de to sammen, ser man at;

1460 var tjenestepiger

35 var væversker

15 var syersker

11 var spindersker

11 var husholdersker

10 var husjomfruer

8 var mejeripiger

7 var skædersker

6 var kokkepiger

4 var bryggerpiger

| ▼ All | ▼ ft | ▼ sogn | ▼ amt | ▼ id | ▼ loknr | ▼ lokalitet | ▼ bygning | ▼ famnr | ▼ fnavn | ▼ enavn | ▼ koen | ▼ famstand | ▼ alder | ▼ civilstand | ▼ giftnr | ▼ erhverv 1 | ▼ erhv |
|-------|------|--------|-------|-------|---------|-------------|-----------|---------|----------------|------------------|--------|---------------|---------|--------------|----------|-------------|--------|
| ★ 🔊 | 10. | 1801 | Alrø | Århus | 10 | 1 | Alrøe Bye | 1 | Else | Jeppesdatter | kvinde | Tjenestepige | 27 | ugift | | | |
| ★ 🔊 | 11. | 1801 | Alrø | Århus | 11 | 1 | Alrøe Bye | 1 | Inger | Jespersdatter | kvinde | Tjenestepige | 23 | ugift | | | |
| ★ 🔊 | 12. | 1801 | Alrø | Århus | 12 | 1 | Alrøe Bye | 1 | Dorthe Sophie | Nielsdatter | kvinde | Tjenestepige | 22 | ugift | | | |
| ★ 🔊 | 13. | 1801 | Alrø | Århus | 13 | 1 | Alrøe Bye | 1 | Inger | Jørgensdatter | kvinde | Tjenestepige | 20 | ugift | | | |
| ★ 🔊 | 24. | 1801 | Alrø | Århus | 24 | 2 | Alrøe Bye | 2 | Anne Kierstine | Simonsdatter | kvinde | Tjenestepige | 28 | ugift | | | |
| ★ 🔊 | 25. | 1801 | Alrø | Århus | 25 | 2 | Alrøe Bye | 2 | Else | Jespersdatter | kvinde | Tjenestepige | 25 | ugift | | | |
| ★ 🔊 | 35. | 1801 | Alrø | Århus | 35 | 3 | Alrøe Bye | 3 | Maren | Sørens datter | kvinde | Tjenestepige | 23 | ugift | | | |
| ★ 🔊 | 49. | 1801 | Alrø | Århus | 49 | 5 | Alrøe Bye | 5 | Maren | Jens datter | kvinde | Datter | 24 | ugift | | | |
| ★ 🔊 | 72. | 1801 | Alrø | Århus | 72 | 8 | Alrøe Bye | 8 | Justine | Wilhelmsdatter | kvinde | Datter | 22 | ugift | | | |
| ★ 🔊 | 81. | 1801 | Alrø | Århus | 81 | 9 | Alrøe Bye | 9 | Anne | Giorderts datter | kvinde | Tjenestepige | 24 | ugift | | | |
| ★ 🔊 | 88. | 1801 | Alrø | Århus | 88 | 10 | Alrøe Bye | 10 | Karen | Jørgensdatter | kvinde | Tjenestepige | 21 | ugift | | | |
| ★ 🔊 | 98. | 1801 | Alrø | Århus | 98 | 11 | Alrøe Bye | 11 | Maren | Michaelsdatter | kvinde | Tjenestepige | 24 | ugift | | | |
| ★ 🔊 | 110. | 1801 | Alrø | Århus | 110 | 13 | Alrøe Bye | 13 | Kiersten | Nielsdatter | kvinde | Datter | 22 | ugift | | | |
| ★ 🔊 | 134. | 1801 | Alrø | Århus | 134 | 16 | Alrøe Bye | 16 | Karen | Michaelsdatter | kvinde | Tjenestepige | 22 | ugift | | | |
| ★ 🔊 | 151. | 1801 | Alrø | Århus | 151 | 18 | Alrøe Bye | 18 | Maren | Sørens datter | kvinde | Tjenestepige | 20 | ugift | | | |
| ★ 🔊 | 163. | 1801 | Alrø | Århus | 163 | 20 | Alrøe Bye | 20 | Maren | Nielsdatter | kvinde | Tjenestepige | 28 | ugift | | | |
| ★ 🔊 | 184. | 1801 | Alrø | Århus | 184 | 27 | Alrøe Bye | 27 | Mette | Nielsdatter | kvinde | hendes datter | 26 | ugift | | Syreske | Halter |
| ★ 🔊 | 195. | 1801 | Alrø | Århus | 195 | 31 | Alrøe Bye | 31 | Cathrine | Sørens datter | kvinde | Datter | 21 | ugift | | | |

Her kan man se at pigerne var tjenestepiger, men dette står under *famstand* da det i 1800-tallet blev tjenestepiger af mange regnet som en del af husstanden i stedet for et erhverv.

Se bilag 1, for eksempel på metode fremgangen