Lea Skriver Hansen
Digital Method for Historians
9/11/2021

Weekly homework 1

Exercise 2:
https://regex101.com/r/1pb666/1

/
(\d{1,2}).(\d{1,2}).+(\d{4})
/
gm

1st Capturing Group

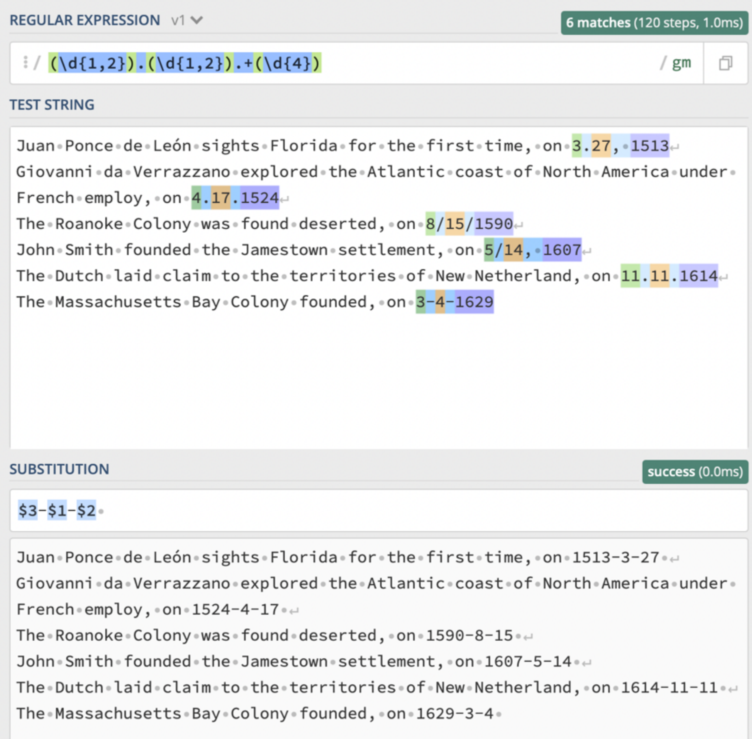
(\d{1,2})
\d
matches a digit (equivalent to [0-9])
{1,2} matches the previous token between 1 and 2times, as many times as possible, giving back as needed (greedy)
. matches any character (except for line terminators)

2nd Capturing Group

(\d{1,2})
\d
matches a digit (equivalent to [0-9])
{1,2} matches the previous token between 1 and 2times, as many times as possible, giving back as needed (greedy)
.
matches any character (except for line terminators)
+ matches the previous token between one and unlimitedtimes, as many times as possible, giving back as needed (greedy)

3rd Capturing Group

(\d{4})
\d
matches a digit (equivalent to [0-9])
{4} matches the previous token exactly 4 times



```
/ (\d{1,2}).(\d{1,2}).+(\d{4}) / gm
```

```
Juan Ponce de León sights Florida for the first time, on 3.27, 1513
Giovanni da Verrazzano explored the Atlantic coast of North America under
French employ, on 4.17.1524
The Roanoke Colony was found deserted, on 8/15/1590
John Smith founded the Jamestown settlement, on 5/14, 1607
The Dutch laid claim to the territories of New Netherland, on 11.11.1614
The Massachusetts Bay Colony founded, on 3-4-1629
```

```
$3-$1-$2
```

```
Juan Ponce de León sights Florida for the first time, on 1513-3-27
Giovanni da Verrazzano explored the Atlantic coast of North America under
French employ, on 1524-4-17
The Roanoke Colony was found deserted, on 1590-8-15
John Smith founded the Jamestown settlement, on 1607-5-14
The Dutch laid claim to the territories of New Netherland, on 1614-11-11
The Massachusetts Bay Colony founded, on 1629-3-4
```

Lea Skriver Hansen
Digital Method for Historians
9/11/2021

Exercise R to Voyant

https://regex101.com/r/2Hqnii/1

```
/
(["])|([,])|([A-Za-zæøåüé0234]+)
/
gm
```

1st Alternative

(["])

1st Capturing Group

(["])

Match a single character present in the list below

["]

" matches the character " with index 34_{10} (22_{16} or 42_{8}) literally (case sensitive)

2nd Alternative

([,])

2nd Capturing Group

([,])

Match a single character present in the list below

[,]

, matches the character , with index 44_{10} ($2C_{16}$ or 54_{8}) literally (case sensitive)

3rd Alternative

([A-Za-zæøåüé0234]+)

3rd Capturing Group

([A-Za-zæøåüé0234]+)

Match a single character present in the list below

[A-Za-zæøåüé0234]

+ matches the previous token between one and unlimited times, as many times as possible, giving back as needed (greedy)

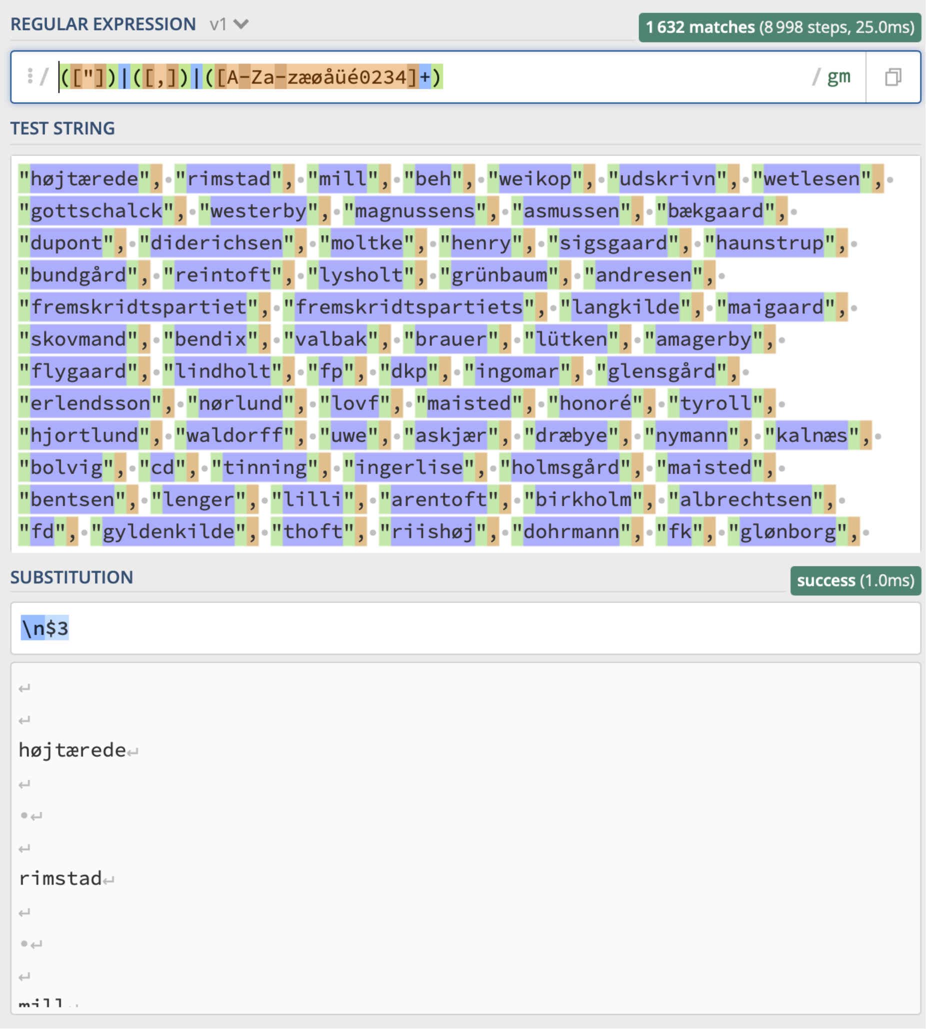
A-Z matches a single character in the range between A (index 65) and Z (index 90) (case sensitive)

a-z matches a single character in the range between a (index 97) and z (index 122) (case sensitive)

æøåüé0234

matches a single character in the list æøåüé0234 (case sensitive)

Lea Skriver Hansen
Digital Method for Historians
9/11/2021



```
/ (["])|([,])|([A-Za-zæøåüé0234]+) / gm
```

```
"højtærede", "rimstad", "mill", "beh", "weikop", "udskrivn", "wetlesen", 
"gottschalck", "westerby", "magnussens", "asmussen", "bækgaard", 
"dupont", "diderichsen", "moltke", "henry", "sigsgaard", "haunstrup", 
"bundgård", "reintoft", "lysholt", "grünbaum", "andresen", 
"fremskridtspartiet", "fremskridtspartiets", "langkilde", "maigaard", 
"skovmand", "bendix", "valbak", "brauer", "lütken", "amagerby", 
"flygaard", "lindholt", "fp", "dkp", "ingomar", "glensgård", 
"erlendsson", "nørlund", "lovf", "maisted", "honoré", "tyroll", 
"hjortlund", "waldorff", "uwe", "askjær", "dræbye", "nymann", "kalnæs", 
"bolvig", "cd", "tinning", "ingerlise", "holmsgård", "maisted", 
"bentsen", "lenger", "lilli", "arentoft", "birkholm", "albrechtsen", 
"fd", "gyldenkilde", "thoft", "riishøj", "dohrmann", "fk", "glønborg", 
```

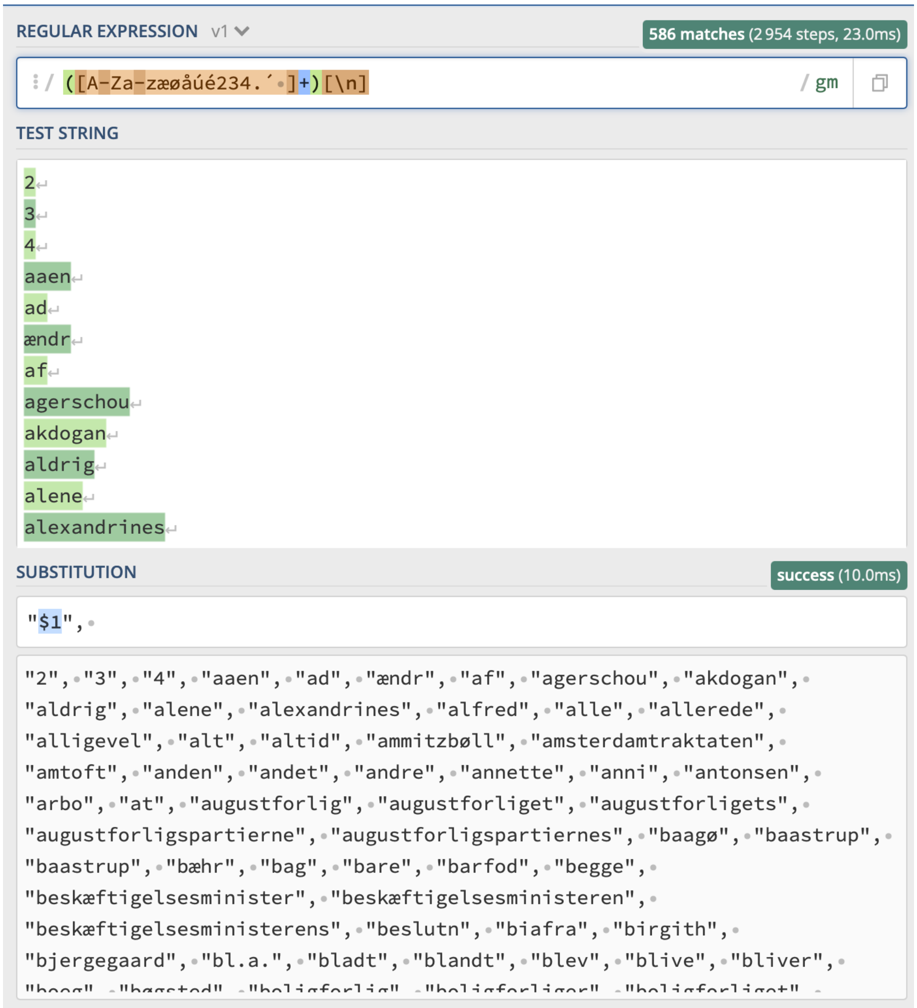
```
\n$3
```

```
↵
↵
højtærede↵
↵
•↵
↵
rimstad↵
↵
•↵
↵
mill↵
```

Exercise Voyant to R:
https://regex101.com/r/0EJ1kg/1
/
([A-Za-zæøåúé234.´]+)[\n]
/
gm
1st Capturing Group
([A-Za-zæøåúé234.´]+)
Match a single character present in the list below
[A-Za-zæøåúé234.´]
+ matches the previous token between one and unlimited times, as many times as possible, giving back as needed (greedy)
A-Z matches a single character in the range between A (index 65) and Z (index 90) (case sensitive)
a-z matches a single character in the range between a (index 97) and z (index 122) (case sensitive)
æøåúé234.´
matches a single character in the list æøåúé234.´ (case sensitive)
Match a single character present in the list below
[\n]

Lea Skriver Hansen
Digital Method for Historians
9/11/2021

\n matches a line-feed (newline) character (ASCII 10)



```
/ ([A-Za-zæøåúé234.´ ]+)[\n] / gm
```

```
2
3
4
aaen
ad
ændr
af
agerschou
akdogan
aldrig
alene
alexandrines
```

```
"$1", 
```

```
"2", "3", "4", "aaen", "ad", "ændr", "af", "agerschou", "akdogan", 
"aldrig", "alene", "alexandrines", "alfred", "alle", "allerede", 
"alligevel", "alt", "altid", "ammitzbøll", "amsterdamtraktaten", 
"amtoft", "anden", "andet", "andre", "annette", "anni", "antonsen", 
"arbo", "at", "augustforlig", "augustforliget", "augustforligets", 
"augustforligspartierne", "augustforligspartiernes", "baagø", "baastrup", 
"baastrup", "bæhr", "bag", "bare", "barfod", "begge", 
"beskæftigelsesminister", "beskæftigelsesministeren", 
"beskæftigelsesministerens", "beslutn", "biafra", "birgith", 
"bjergegaard", "bl.a.", "bladt", "blandt", "blev", "blive", "bliver", 
"boeg", "bøgsted", "boligforlig", "boligforliger", "boligforliget", 
```

Spreadsheet Quistion:

"What are the basic principles for using spreadsheets for good data organisation?"

There are a few steps along the way that you can do to make spreadsheets good for working with data. The first advice to make data good to work with in a spreadsheet is to be consistent, when working in a spreadsheet it is important for the data that you use the word for things in the entire spreadsheet to get the best and most consistent result of your data. So, when your data is f.ex. about gender use the word female only, and not write woman if you do the data con get a bit confusing. The second good thing to do is, especially if your data contains a lot of dates, is to write the dates in the universal form that fits most programs. So, if you're working with the date of danish constitution you should write the date like this: 1849-06-05, this way every spreadsheet program you are working with will understand it as a date. Third, it is important to only put one piece of data in a cell and not multiple. This is so the data don't get mixed wrong, and you get a result which is useless. On this note it is also important that there are no empty cells in your spreadsheets, this is to prevent the spreadsheet to think that data is missing, and it messes up with your result

And lastly it is important to save your data as you're working and make backups more than one place, maybe on your computer and on a hard disk, so your data doesn't get lost.