W45_Assignment_Mathias_Thomasen_Madsen

1. Create a *tidy* spreadsheet/table listing the names of Danish monarchs with their birth- and death-date and duration of reign. They should be sortable by year of birth. Suitable source websites are here and here, but you can also use another source, provided you reference it.

As my source I've used:

https://www.kongehuset.dk/monarkiet-i-danmark/kongerakken

The spreadsheet can be found in my github:

https://github.com/Digital-Methods-HASS/au681088_Madsen_Thomasen_Mathias

I haven't included civil wars or other periods where the kingdom was split between multiple kings, or there wasn't a king.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	duration as regent start	duration as regent end	certain?	name	birth year	birth month	birth day	Certain?	death year	death month	death day	certain?	years as regent
2	936	958	no	Gorm den Gamle	908	NA	NA	no	958	NA	NA	no	22
3	958	987	no	Harald 1. Blåtand	NA	NA	NA	no	987	NA	NA	no	29
4	987	1014	no	Svend 1. Tveskæg	NA	NA	NA	no	1014	NA	NA	yes	27
5	1014	1018	yes	Harald 2.	NA	NA	NA	no	1018	NA	NA	no	4
6	1018	1035	yes	Knud 1. den Store	995	NA	NA	no	1035	NA	NA	yes	17
7	1035	1042	yes	Hardeknud	1020	NA	NA	yes	1042	NA	NA	yes	7
8	1042	1047	yes	Magnus den Gode	1024	NA	NA	yes	1047	NA	NA	yes	5
9	1047	1074	yes	Svend 2. Estridsen	NA	NA	NA	no	1076	april	28	yes	27
10	1074	1080	yes	Harald 3. Hen	NA	NA	NA	no	1080	NA	NA	yes	6
11	1080	1086	yes	Knud 2. den Hellige	NA	NA	NA	no	1086	july	10	yes	6
12	1086	1095	yes	Oluf 1. Hunger	NA	NA	NA	no	1095	NA	NA	yes	9
13	1095	1103	yes	Erik 1. Ejegod	1056	NA	NA	yes	1103	NA	NA	yes	8
14	1104	1134	yes	Niels	NA	NA	NA	no	1134	NA	NA	yes	30
15	1134	1137	yes	Erik 2. Emune	NA	NA	NA	no	1137	july	18	yes	3
16	1137	1146	yes	Erik 3. Lam	NA	NA	NA	no	1146	august	27	yes	9
17	1157	1182	yes	Valdemar 1. den Store	1131	NA	NA	yes	1182	NA	NA	yes	25
18	1182	1202	yes	Knud 4.	1163	NA	NA	yes	1202	NA	NA	yes	20
19	1202	1241	yes	Valdemar 2. Sejre	1170	NA	NA	yes	1241	NA	NA	yes	39
20	1241	1250	yes	Erik 4. Plovpenning	1216	NA	NA	yes	1250	NA	NA	yes	9
21	1250	1252	yes	Abel	1218	NA	NA	no	1252	NA	NA	yes	2
22	1252	1259	yes	Christoffer 1.	1219	NA	NA	no	1259	NA	Na	yes	7
23	1259	1286	yes	Erik 5. Kippling	1249	NA	NA	yes	1286	NA	NA	yes	27
24	1286	1319	yes	Erik 6. Menved	1274	NA	NA	yes	1319	NA	NA	yes	33
25	1319	1326	yes	Christoffer 2.	1276	NA	NA	yes	1332	NA	Na	yes	10
26	1326	1329	yes	Valdemar 3.	1315	NA	NA	no	1364	NA	NA	yes	3
27	1329	1332	yes	Christoffer 2.	1276	NA	NA	yes	1332	NA	NA	yes	10
28	1340	1375	yes	Valdemar 4. Atterdag	1320	NA	NA	no	1375	NA	NA	yes	35
29	1375	1387	yes	Oluf 2.	1370	NA	NA	yes	1387	NA	NA	yes	12
30	1387	1396	yes	Margrete 1.	1353	NA	NA	yes	1412	october	28	yes	9
31	1396	1439	yes	Erik 7. af Pommern	1382	NA	NA	no	1459	NA	NA	yes	43
32	1440	1444	yes	Christoffer 3. af Bayern	1416	NA	NA	yes	1448	NA	NA	yes	4
33	1448	1481	yes	Christian 1.	1426	NA	NA	yes	1481	may	21	yes	33
34	1482	1513	yes	Hans	1455	2	2	yes	1513	february	20	yes	31
35	1513	1523	yes	Christian 2.	1481	7	1	yes	1559	january	25	yes	10
36	1523	1533	yes	Frederik 1.	1471	10	7	yes	1533	april	10	yes	10
37	1536	1559	yes	Christian 3.	1503	8	12	yes	1559	january	1	yes	23
38	1559	1588	yes	Frederik 2.	1534	7	1	yes	1588	april	4	yes	29
39	1588	1648	yes	Christian 4.	1577	4	12	yes	1648	february	28	yes	60
40	1648	1670	yes	Frederik 3.	1609	3	18	yes	1670	february	9	yes	22
41	1670	1699	yes	Christian 5.	1646	4	15	yes	1699	august	25	yes	29
42	1699	1730	yes	Frederik 4.	1671	10	11	yes	1730	october	12	yes	31
43	1730	1746	yes	Christian 6.	1699	11	30	yes	1746	august	6	yes	16
44	1746	1766	yes	Frederik 5.	1723	3	31	yes	1766	january	14	yes	20
45	1766	1808	yes	Christian 7.	1749	1	29	yes	1808	march	13	yes	42
46	1808	1839	yes	Frederik 6.	1768	1	28	yes	1839	december	3	yes	31
47	1839	1848	yes	Christian 8.	1786	9	18	yes	1848	january	20	yes	9
48	1848	1863	yes	Frederik 7.	1808	10	6	yes	1863	november	15	yes	15
49	1863	1906	yes	Christian 9	1818	4	8	yes	1906	january	29	yes	43
50	1906	1912	yes	Frederik 8.	1843	6	3	yes	1912	may	14	yes	6
51	1912	1947	yes	Christian 10.	1870	9	26	yes	1947	april	20	yes	35
52	1947	1972	yes	Frederik 9.	1899	3	11	yes	1972	january	14	yes	25
53	1972	NA	yes	Margrete 2.	1940	4	16	yes	NA	NA	NA	NA	NA

Thoughts behind the spreadsheet:

- I wanted to follow the steps regarding how to make a proper spreadsheet, using the steps given in the article: Data Organization in Spreadsheets, Karl W. Broman, Kara H. Woo. https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1375989
- I especially tried to only have singular values in each column and have no empty columns.
- I wanted to do the spreadsheet in a format where I used the YYYY-MM-DD format.
- I found I couldn't use a 0 in front of a number. From google I found that you could format a singular cell. I tried to format the singular cell, but every time I tried, excel would freeze and stop responding. As a solution I just typed the name of the month or wrote it as a single digit number. Later this can be changed in OpenRefine.

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

No it does not. What it does is that it organizes it through the preview, where you have the chance to alternate the layout. Afterwards it allows you to manually change the spreadsheet and use other tools for analyses.

3. Fix the interviews dataset in OpenRefine enough to answer this question: "Which two months are reported as the most water-deprived/driest by the interviewed farmer households?"

The column named "months_no_water must necessarily be the column that describes the months that had no water available. I opened the text facet and got 14 choices. To simplify the data, I tried to cluster it. It only gave me the option to merge some of the data, but doing so it would also remove some of the answers being December. Hence I calculated all the times the month December was mentioned before merging in the cluster. After having merged some of the data in the cluster, I couldn't find another way to furthermore ease the calculation of how many times each month was mentioned in the column. So I calculated them all manually and got the results:

October = 74

September = 70

November = 37

August = 33

December = 11

January = 2

July = 2

May = 1

June = 1

April = 1

As my results suggest, October and September where the two months mentioned the most, so they must necessarily be the two most water-deprived months.