Data Fundamentals Lab 6: Exploring Data in Larger Data Sets

In this lab, you will use a spreadsheet tool to prepare, filter, sort and transpose data from the World Health Organisation about main infectious diseases filtered by countries and years. Countries selected are [UNDP Africa](#)[1].

Hypothesis

Vaccine preventable infectious diseases continue to plague countries in the region and impact overall health of citizens.

Questions

- Q1: What is the most common selected vaccine preventable infectious disease in the region from 2009 to 2014?
- Q2: What are the three most common selected vaccine preventable infectious disease in the region in 2014?
- Q3: What are the most common selected vaccine preventable infectious disease in Tanzania?
- Q4: What are the most common selected vaccine preventable infectious diseases in your country?
- Q5: What countries have the highest rate of all these diseases in 2014? What data would you need to figure out which country has the highest rate of infection?

---

[1] South Sudan is not included because WHO did not have data about it.

- Q6: Are the number of infections by the most common selected vaccine preventable infectious diseases going up or down?

Task 1: Get Data

1. Open **Selected Vaccine Preventable Infectious diseases.xlsx** in your lab folder
2. Review all the sheets to make sure you understand what each contains and is being measured.
   a. **Source:** First recommendation is to make a sheet with the source of your dataset. Sometime if there is some special definitions of your data you include that in source sheet.
   b. **Data**: This is the data directly extracted from your source without any change.
   c. **Copy:** Make a copy of the original data sheet in case you make a mistake
   d. **Clean:** Is a cleaned sheet to be used for analysis.
   e. **SubSaharanAfrica:** Is a filtered sheet with just the Sub Saharan Africa countries to analyse.
   f. **PT_SubSaharanAfrica:** PivotTable (we will explore in the next lab)
   g. **Total by Country:** Total reported cases from 2009 to 2014

Task 2: Analyse Data

Q1: What is the most common selected vaccine preventable infectious disease in the region from 2009 to 2014?

3. Go to **Total by Country** sheet

4. Copy data. Do not use copy all (Control A), just select data and copy it

5. Create a new sheet and name it Q1.

6. Go to Home, select paste and then **Transpose**.

7. You will notice that columns are now in rows and rows are in columns. Delete A1 and rename as **Diseases**.

A1:AMJ1 — Diseases

Diseases	Angola	Benin	Botswana	Burkina Faso	Burundi	Cabo Verde	Cameroon	Central Africa	Chad	Comoros	Congo	Cote d'Ivoire	Democratic	Equatorial G	Eritrea	Ethiopia
SUM of Num	786	0	0	0	0	0	75	2218	1503	0	315	48	33243	21	0	0
SUM of Num	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SUM of Num	13396	3797	15	163	4113	0	37743	138	23729	0	3660	2013	141291	0	0	33191
SUM of Cong	0	0	0	0	0	62	0	0	0	0	0	0	1	0	0	0
SUM of Num	0	3717	0	28680	0	0	7092	1772	15128	0	0	1522	47883	0	0	2237
SUM of Diph	35	0	0	0	0	0	0	35	0	0	1	0	0	0	20	0
SUM of Japa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SUM of Lepr	3802	1410	5	1506	1271	8	2797	881	1522	1624	344	4763	24683	54	12	26035
SUM of Mala	0	0	1081	0	168	0	0	0	0	0	0	0	0	0	0	0
SUM of Mea	30126	3990	1054	65569	978	0	3195	1639	10630	5	775	1487	333387	1602	547	31005
SUM of Mum	0	0	0	8090	0	295	0	0	0	109	0	0	0	0	10170	0
SUM of Neon	212	28	0	14	20	0	198	241	1195	5	13	50	4062	5	0	234
SUM of Pert	6479	53	0	97	0	25	0	653	0	0	170	0	8846	0	310	0
SUM of Plag	0	0	0	0	0	0	0	0	0	0	0	0	283	0	0	0
SUM of Polio	38	0	0	0	0	0	13	4	167	0	442	36	193	5	0	10
SUM of Rube	285	95	1020	761	708	11388	629	611	6	36	195	575	4986	0	732	2105
SUM of Total	3351	110	0	208	20	4	424	250	916	6	25	44	4846	5	101	174
SUM of Tube	0	0	0	0	0	2064	0	0	0	634	0	0	0	2410	0	0
SUM of Yello	0	0	0	1032	0	0	102	46	257	0	191	113	488	0	0	223
SUM of Num	0	0	0	3470	0	0	0	0	0	0	0	0	0	0	0	0
SUM of Num	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9651

8. In **AV1** type **Total Cases**

9. Use the **SUM** function to calculate the total number of cases

AN	AO	AP	AQ	AR	AS	AT	AU	AV
South Africa	South Sudan	Swaziland	Togo	Uganda	United Repub	Zambia	Zimbabwe	Total Cases
0	1341	0	0	282	0	0	0	40313
0	0	0	0	506	26	39	22	772
10521	6421	19	783	10819	13667	12034	70347	670834
1	0	0	3	0	0	10	65	149
0	370	0	2275	0	0	0	0	204743
1	0	0	0	0	0	0	0	228
0	0	0	0	0	0	0	0	0
15	4867	0	259	1569	11483	1196	15	135940
0	0	1019	0	0	0	0	0	5123
18571	4174	339	2111	16445	5304	29954	10549	831004
0	0	14378	0	0	0	0	0	36870
7	118	0	137	1095	13	0	22	9695
531	0	6	259	0	0	681	0	63205
0	0	0	0	0	16	0	0	868
0	2	0	0	4	0	0	0	1276
8652	20	385	323	13508	839	1528	1755	67115
4	32	2	137	9054	13	0	22	26056
0	0	0	0	0	0	0	0	6708
0	128	0	12	204	0	0	0	3909
0	0	0	0	0	0	0	0	3694
0	13437	0	0	930	0	0	0	25204

10. Q1 is referring to total cases so let's hide countries for a better view.

Select column B to AU, use right click on any column and click on **Hide**.

	A	AV
1	Diseases	Total Cases
2	SUM of Number of new reported cases of human African trypanosomiasis (T.b. gambiense)	40313
3	SUM of Number of new reported cases of human African trypanosomiasis (T.b. rhodesiense)	772
4	SUM of Number of reported cases of cholera	670834
5	SUM of Congenital Rubella Syndrome - number of reported cases	149
6	SUM of Number of suspected meningitis cases reported	204743
7	SUM of Diphtheria - number of reported cases	228
8	SUM of Japanese encephalitis - number of reported cases	0
9	SUM of Leprosy - number of reported cases	135940
10	SUM of Malaria - number of reported confirmed cases	5123
11	SUM of Measles - number of reported cases	831004
12	SUM of Mumps - number of reported cases	36870
13	SUM of Neonatal tetanus - number of reported cases	9695
14	SUM of Pertussis - number of reported cases	63205
15	SUM of Plague - number of reported cases	868
16	SUM of Poliomyelitis - number of reported cases	1276
17	SUM of Rubella - number of reported cases	67115
18	SUM of Total tetanus - number of reported cases	26056
19	SUM of Tuberculosis - new and relapse cases	6708
20	SUM of Yellow fever - number of reported cases	3909
21	SUM of Number of cases of cutaneous leishmaniasis reported	3694
22	SUM of Number of cases of visceral leishmaniasis reported	25204
23		

11. Select row 1 and use the filter function.

	A	AV
1	Diseases	Total Cases
2	SUM of Number of new reported cases of human African trypanosomiasis (T.b. gambiense)	40313
3	SUM of Number of new reported cases of human African trypanosomiasis (T.b. rhodesiense)	772
4	SUM of Number of reported cases of cholera	670834
5	SUM of Congenital Rubella Syndrome - number of reported cases	149
6	SUM of Number of suspected meningitis cases reported	204743
7	SUM of Diphtheria - number of reported cases	228
8	SUM of Japanese encephalitis - number of reported cases	0
9	SUM of Leprosy - number of reported cases	135940
10	SUM of Malaria - number of reported confirmed cases	5123
11	SUM of Measles - number of reported cases	831004
12	SUM of Mumps - number of reported cases	36870
13	SUM of Neonatal tetanus - number of reported cases	9695
14	SUM of Pertussis - number of reported cases	63205
15	SUM of Plague - number of reported cases	868
16	SUM of Poliomyelitis - number of reported cases	1276
17	SUM of Rubella - number of reported cases	67115
18	SUM of Total tetanus - number of reported cases	26056
19	SUM of Tuberculosis - new and relapse cases	6708
20	SUM of Yellow fever - number of reported cases	3909
21	SUM of Number of cases of cutaneous leishmaniasis reported	3694
22	SUM of Number of cases of visceral leishmaniasis reported	25204

12. Sort Descending Total Cases column.

	A	AV
1	Diseases	Total Cases
2	SUM of Measles - number of reported cases	831004
3	SUM of Number of reported cases of cholera	670834
4	SUM of Number of suspected meningitis cases reported	204743
5	SUM of Leprosy - number of reported cases	135940
6	SUM of Rubella - number of reported cases	67115
7	SUM of Pertussis - number of reported cases	63205
8	SUM of Number of new reported cases of human African trypanosomiasis (T.b. gambiense)	40313
9	SUM of Mumps - number of reported cases	36870
10	SUM of Total tetanus - number of reported cases	26056
11	SUM of Number of cases of visceral leishmaniasis reported	25204
12	SUM of Neonatal tetanus - number of reported cases	9695
13	SUM of Tuberculosis - new and relapse cases	6708
14	SUM of Malaria - number of reported confirmed cases	5123
15	SUM of Yellow fever - number of reported cases	3909
16	SUM of Number of cases of cutaneous leishmaniasis reported	3694
17	SUM of Poliomyelitis - number of reported cases	1276
18	SUM of Plague - number of reported cases	868
19	SUM of Number of new reported cases of human African trypanosomiasis (T.b. rhodesiense)	772
20	SUM of Diphtheria - number of reported cases	228
21	SUM of Congenital Rubella Syndrome - number of reported cases	149
22	SUM of Japanese encephalitis - number of reported cases	0

Q1: What is the most common selected vaccine preventable infectious disease in the region from 2009 to 2014?

Answer: Measles with a total of 831,004 reported cases

Q2: What are the three most common selected vaccine preventable infectious diseases in the region in 2014?

13. Go to SubSaharanAfrica sheet.
14. Filter Year (Column B) and leave just 2014

15. Copy the filtered data.

16. Create a new sheet and name it Q2

17. Go to Home, select paste and then **Paste Special**. In the window select **Values** and **Transpose**. Click on **OK**.

18. Delete row 2 (Year)

19. Type in AV1 **Total Cases 2014**

20. Use the SUM function to calculate total of cases

	AN	AO	AP	AQ	AR	AS	AT	AU	AV
1	South Africa	South Sudan	Swaziland	Togo	Uganda	United Repu‣	Zambia	Zimbabwe	Total Cases 2014
2		63		0	9				3679
3			No data		69	1	12	3	117
4		6421		262	309				102467
5		111		351					21434
6	1	0		0			10	2	14
7		0	0	0				0	1
8	0	0							0
9	No data	691	No data	97	275	1947	No data	0	18597
10	11 705		269	1 130 251	3 631 939	680 807	4 077 547	535 931	295
11	66	441	0	577	314	88	9	0	73914
12		0	3						7
13	2	0	0	18	84	2		0	835
14	230	0	0	38				0	10098
15									0
16	0	2	0	0	0	0	0	0	49
17	10	0	1	16	291	529	74	1130	7399
18	2	0	0	18	1358	2		0	2900
19	306 166	8 335	5 583	2 525	44 187	61 571	37 931	29 653	719
20	0	0		0				0	31
21									0
22									0

21. Select row 1 and use the Filter function

22. Sort the **Total Cases 2014** column by Sort Descending (or Smallest to Largest in Excel**).**

	A	AV
1	Country	Total Cases 201
2	Number of reported cases of cholera	102467
3	Measles - number of reported cases	73914
4	Number of suspected meningitis cases reported	21434
5	Leprosy - number of reported cases	18597
6	Pertussis - number of reported cases	10098
7	Rubella - number of reported cases	7399
8	Number of new reported cases of human African trypanosomiasis (T.b. gambiense)	3679
9	Total tetanus - number of reported cases	2900
10	Neonatal tetanus - number of reported cases	835
11	Tuberculosis - new and relapse cases	719
12	Malaria - number of reported confirmed cases	295
13	Number of new reported cases of human African trypanosomiasis (T.b. rhodesiense)	117
14	Poliomyelitis - number of reported cases	49
15	Yellow fever - number of reported cases	31
16	Congenital Rubella Syndrome - number of reported cases	14
17	Mumps - number of reported cases	7
18	Diphtheria - number of reported cases	1
19	Japanese encephalitis - number of reported cases	0
20	Plague - number of reported cases	0
21	Number of cases of cutaneous leishmaniasis reported	0
22	Number of cases of visceral leishmaniasis reported	0

Q2: What are the three most common selected vaccine preventable infectious diseases in the region in 2014?

Answer: Cholera (102,467), measles (73,914) and meningitis (21,434) were the 3 most common in 2014.

Q3: What are the most common selected vaccine preventable infectious disease in Tanzania?

23. Make a new sheet and name it Q3

24. Go to SubSaharanAfrica sheet. In column A (Country) filter just **United Republic of Tanzania**.

25. Copy the filtered data

26. In Q3, paste special with values and transpose.

	A	B	C	D	E	F	G
1	Country	United Rep	United Rep	United Rep	United Rep	United Rep	United Repub
2	Year	2014	2013	2012	2011	2010	2009
3	Number of r						
4	Number of r	1	1	4	1	5	14
5	Number of r		270	286	942	4469	7700
6	Number of s						
7	Congenital I						
8	Diphtheria -		0	0	0	0	0
9	Japanese e		0	0			
10	Leprosy - nu	1947	2005	2528	No data	2349	2654
11	Malaria - nu	680 807	1 552 444	1 986 955	2 150 761	1 278 998	3 242
12	Measles - n	88	185	1668	1622	167	1574
13	Mumps - nu						
14	Neonatal tet	2	0	0	0	5	6
15	Pertussis - r		0	0	0	0	0
16	Plague - nur			7		9	
17	Poliomyelitis	0	0	0	0	0	
18	Rubella - nu	529	116	55	18	121	0
19	Total tetanu	2				5	6
20	Tuberculosi	61 571	64 053	62 479	59 357	61 098	61 537
21	Yellow fever		0	0			
22	Number of c						
23	Number of c						

27. Delete Row 1 and rename Column A as **Diseases**.

28. Name Column H as **Cases of Infectious Diseases in Tanzania**.

	A	B	C	D	E	F	G	H
1	Diseases	2014	2013	2012	2011	2010	2009	Cases of Infectious Diseases in Tanzania
2	Number of r							
3	Number of r	1	1	4	1	5	14	
4	Number of r		270	286	942	4469	7700	
5	Number of s							
6	Congenital I							
7	Diphtheria -		0	0	0	0	0	
8	Japanese er		0	0				
9	Leprosy - nu	1947	2005	2528	No data	2349	2654	
10	Malaria - nu	680 807	1 552 444	1 986 955	2 150 761	1 278 998	3 242	
11	Measles - n	88	185	1668	1622	167	1574	
12	Mumps - nu							
13	Neonatal tet	2	0	0	0	5	6	
14	Pertussis - r		0	0	0	0	0	
15	Plague - nur			7		9		
16	Poliomyelitis	0	0	0	0	0		
17	Rubella - nu	529	116	55	18	121	0	
18	Total tetanu:	2				5	6	
19	Tuberculosis	61 571	64 053	62 479	59 357	61 098	61 537	
20	Yellow fever		0	0				
21	Number of c							
22	Number of c							

29. In H3, use SUM function to calculate Cases of Infectious Diseases from 2009 to 2014.

	A	B	C	D	E	F	G	H
1	Diseases	2014	2013	2012	2011	2010	2009	Cases of Infectious Diseases in Tanzania
2	Number of r							0
3	Number of r	1	1	4	1	5	14	26
4	Number of r		270	286	942	4469	7700	13667
5	Number of s							0
6	Congenital I							0
7	Diphtheria -		0	0	0	0	0	0
8	Japanese er		0	0				0
9	Leprosy - nu	1947	2005	2528	No data	2349	2654	11483
10	Malaria - nu	680 807	1 552 444	1 986 955	2 150 761	1 278 998	3 242	0
11	Measles - n	88	185	1668	1622	167	1574	5304
12	Mumps - nu							0
13	Neonatal tet	2	0	0	0	5	6	13
14	Pertussis - r		0	0	0	0	0	0
15	Plague - nur			7		9		16
16	Poliomyelitis	0	0	0	0	0		0
17	Rubella - nu	529	116	55	18	121	0	839
18	Total tetanu:	2				5	6	13
19	Tuberculosis	61 571	64 053	62 479	59 357	61 098	61 537	0
20	Yellow fever		0	0				0
21	Number of c							0
22	Number of c							0

30. Hide Columns B to G

31. Use the filter function to sort the largest to the smallest column H.

	A	H
1	Diseases	Cases of Infectious Diseases in Tanzania
2	Number of reported cases of cholera	13667
3	Leprosy - number of reported cases	11483
4	Measles - number of reported cases	5304
5	Rubella - number of reported cases	839
6	Number of new reported cases of human African trypanosomiasis (T.b. rhodesiense)	26
7	Plague - number of reported cases	16
8	Neonatal tetanus - number of reported cases	13
9	Total tetanus - number of reported cases	13
10	Number of new reported cases of human African trypanosomiasis (T.b. gambiense)	0
11	Number of suspected meningitis cases reported	0
12	Congenital Rubella Syndrome - number of reported cases	0
13	Diphtheria - number of reported cases	0
14	Japanese encephalitis - number of reported cases	0
15	Malaria - number of reported confirmed cases	0
16	Mumps - number of reported cases	0
17	Pertussis - number of reported cases	0
18	Poliomyelitis - number of reported cases	0
19	Tuberculosis - new and relapse cases	0
20	Yellow fever - number of reported cases	0
21	Number of cases of cutaneous leishmaniasis reported	0
22	Number of cases of visceral leishmaniasis reported	0
23		
24		

32. Q3: What are the most common selected vaccine preventable infectious diseases in Tanzania?

Answer:

Diseases	Cases of Infectious Diseases in Tanzania
Number of reported cases of cholera	13667
Leprosy - number of reported cases	11483
Measles - number of reported cases	5304
Rubella - number of reported cases	839