

Data aggregations and descriptive statistics

Summarising data

Data overview

You are an analyst hired by an NPO to work on initiatives that focus on closing the **gender achievement gap in education**.



You need...

Summary statistics to help you understand the gender differences among PhD graduates.



The dataset

The **PhD graduates** in public chartered universities in Kenya, 2015 dataset from openAFRICA.

Data overview

The PhD graduates dataset contains 22 rows and the following columns:

A. Qualification

The title of the qualification being considered.

B. Male

The number of male graduates for each qualification in the year 2015.

C. Female

The number of female graduates for each qualification in the year 2015.



The dataset

	A	B	C	
1	Qualification	Male	Female	
2	Agriculture, Forestry and Fisheries	17	13	
3	Architecture	1	1	
4	Business and administration	69	38	
5	Computing	6	3	
6	Education (Arts)	39	21	
7	Education (Science)	4	0	
8	Engineering	2	0	
9	Environment	8	8	
10	Health and Welfare	26	18	
11	Humanities and Arts	36	21	
12	Journalism and Information	7	2	
13	Law	1	1	
14	Life Science and Physical Science	25	15	
15	Manufacturing	0	0	
16	Mathematics and Statistics	7	1	
17	Security and Conflict Resolution	0	0	
18	Services	2	2	
19	Social and Behavioral Science	5	6	
20	Teacher Training	0	2	
21	Veterinary	7	3	
22	Other	0	0	

Consider the questions we want to investigate



PhD graduates
in public chartered universities in
Kenya

How many males and
females **received PhD**
qualifications, respectively?

How many qualifications
are taken into account?

Which qualifications have the
least or **most** male and
female **graduates**,
respectively?

The SUM function

How many males and females received PhD qualifications, respectively?

The answer to this question will give more information on the **gender disparities** among PhD holders at Kenya's public chartered universities in 2015.

01. Add all the values in the **Male** column;

02. Add all the values in the **Female** column;

	A	B	C	D
1	Qualification	Male	Female	
2	Agriculture, Forestry and Fisheries	17	13	
3	Architecture	1	1	
4	Business and administration	69	38	
5	Computing	6	3	
6	Education (Arts)	39	21	
7	Education (Science)	4	0	
8	Engineering	2	0	
9	Environment	8	8	
10	Health and Welfare	26	18	
11	Humanities and Arts	36	21	
12	Journalism and Information	7	2	
13	Law	1	1	
14	Life Science and Physical Science	25	15	
15	Manufacturing	0	0	
16	Mathematics and Statistics	7	1	
17	Security and Conflict Resolution	0	0	
18	Services	2	2	
19	Social and Behavioral Science	5	6	
20	Teacher Training	0	2	
21	Veterinary	7	3	
22	Other	0	0	

The SUM function

=SUM(value1, [value2, ...])

The **SUM** function is used to **add** the cells that are specified in the function argument.

13	Law	1
14	Life Science and Physical Science	25
15	Manufacturing	0
16	Mathematics and Statistics	7
17	Security and Conflict Resolution	0
18	Services	2
19	Social and Behavioral Science	5
20	Teacher Training	0
21	Veterinary	7
22	Other	0
23	Total	=SUM(B2:B22)
24		

The SUM of a range.

13	Law	1	1
14	Life Science and Physical Science	25	15
15	Manufacturing	0	0
16	Mathematics and Statistics	7	1
17	Security and Conflict Resolution	0	0
18	Services	2	2
19	Social and Behavioral Science	5	6
20	Teacher Training	0	2
21	Veterinary	7	3
22	Other	0	0
23	Total	=SUM(B14,B11,B6,B10)	
24			

The SUM of specific cells.

The SUM function

- 01.** Ignores **empty** cells, cells with **text**, and **True/False** values.
- 02.** Returns an **error** if any of the cells contain errors.

14	Life Science and Physical Science	25	15
15	Manufacturing	0	0
16	Mathematics and Statistics	7	1
17	Security and Conflict Resolution	0	0
18	Services	2	2
19	Social and Behavioral Science	5	6
20	Teacher Training	one	2
21	Veterinary	TRUE	3
22	Other		0
23	Total	255	
24			

13	Law	1	1
14	Life Science and Physical Science	25	15
15	Manufacturing	0	0
16	Mathematics and Statistics	7	1
17	Security and Conflict Resolution	0	0
18	Services	2	2
19	Social and Behavioral Science	5	6
20	Teacher Training	#NAME?	2
21	Veterinary	7	3
22	Other	0	0
23	Total	#NAME?	
24			

The COUNT function

How many qualifications are taken into account?

This question will give insight into the **number of qualifications** that are taken into account in this study and give us a better understanding of the **scope of the values** we are working with.

01.

Count the number of entries in the **Qualification** column;

	A	B	C	D
1	Qualification	Male	Female	
2	Agriculture, Forestry and Fisheries	17	13	
3	Architecture	1	1	
4	Business and administration	69	38	
5	Computing	6	3	
6	Education (Arts)	39	21	
7	Education (Science)	4	0	
8	Engineering	2	0	
9	Environment	8	8	
10	Health and Welfare	26	18	
11	Humanities and Arts	36	21	
12	Journalism and Information	7	2	
13	Law	1	1	
14	Life Science and Physical Science	25	15	
15	Manufacturing	0	0	
16	Mathematics and Statistics	7	1	
17	Security and Conflict Resolution	0	0	
18	Services	2	2	
19	Social and Behavioral Science	5	6	
20	Teacher Training	0	2	
21	Veterinary	7	3	
22	Other	0	0	

The COUNT function

=COUNT(value1, [value2, ...])

The **COUNT** function counts the number of cells that have numerical values within the specified range.

15	Manufacturing	0
16	Mathematics and Statistics	7
17	Security and Conflict Resolution	0
18	Services	2
19	Social and Behavioral Science	5
20	Teacher Training	0
21	Veterinary	7
22	Other	0
23	Total	262
24	Number of Qualification	=COUNT(B2:B22)
25		
26		

The COUNT of a range.

The COUNT function

01. Ignores **empty** cells, cells with **text**, and **True/False** values.

02. Use **COUNTA** to include text and True/False values.

03. Ignores cells that contain **errors**.

16	Mathematics and Statistics	7	1
17	Security and Conflict Resolution	0	0
18	Services	2	2
19	Social and Behavioral Science	5	6
20	Teacher Training	one	2
21	Veterinary		3
22	Other	TRUE	0
23	Total	255	
24	Number of Qualifications	18	
25			
26			
27			

15	Manufacturing	0	0
16	Mathematics and Statistics	7	1
17	Security and Conflict Resolution	0	0
18	Services	2	2
19	Social and Behavioral Science	5	6
20	Teacher Training	one	2
21	Veterinary	7	3
22	Other	TRUE	0
23	Total	202	
24	Number of Qualifications	=COUNTA(B2:B22)	
25			
26			

15	Manufacturing	0	0
16	Mathematics and Statistics	7	1
17	Security and Conflict Resolution	0	0
18	Services	2	2
19	Social and Behavioral Science	5	6
20	Teacher Training	#NAME?	2
21	Veterinary	7	3
22	Other	0	0
23	Total	#NAME?	
24	Number of Qualifications	21	
25			
26			

The MIN and MAX functions

Which qualifications have the least or most male and female graduates, respectively?

This inquiry will shed light on the **degrees that have graduated the most or least** number of **males** and **females**.

This can also help us determine the qualifications that men and women are more likely to pursue.

01. Find the minimum and maximum values in the **Male** column;
02. Find the minimum and maximum values in the **Female** column;

	A	B	C	D
1	Qualification	Male	Female	
2	Agriculture, Forestry and Fisheries	17	13	
3	Architecture	1	1	
4	Business and administration	69	38	
5	Computing	6	3	
6	Education (Arts)	39	21	
7	Education (Science)	4	0	
8	Engineering	2	0	
9	Environment	8	8	
10	Health and Welfare	26	18	
11	Humanities and Arts	36	21	
12	Journalism and Information	7	2	
13	Law	1	1	
14	Life Science and Physical Science	25	15	
15	Manufacturing	0	0	
16	Mathematics and Statistics	7	1	
17	Security and Conflict Resolution	0	0	
18	Services	2	2	
19	Social and Behavioral Science	5	6	
20	Teacher Training	0	2	
21	Veterinary	7	3	
22	Other	0	0	

The MIN and MAX functions

=MIN(value1, [value2, ...])

The **MIN** and **MAX** functions find the minimum and maximum number within the specified range.

=MAX(value1, [value2, ...])

15	Manufacturing	0
16	Mathematics and Statistics	7
17	Security and Conflict Resolution	0
18	Services	2
19	Social and Behavioral Science	5
20	Teacher Training	0
21	Veterinary	7
22	Other	0
23	Total	262
24	Number of Qualifications	21
25	Minimum	=MIN(B2:B22)
26		

The MIN of a range.

16	Mathematics and Statistics	7
17	Security and Conflict Resolution	0
18	Services	2
19	Social and Behavioral Science	5
20	Teacher Training	0
21	Veterinary	7
22	Other	0
23	Total	262
24	Number of Qualifications	21
25	Minimum	0
26	Maximum	=MAX(B2:B22)
27		

The MAX of a range.

The MIN and MAX functions

01. Ignores **empty** cells, cells with **text**, and **True/False** values.

02. Returns an **error** if any of the cells contain errors.

16	Mathematics and Statistics	7	1
17	Security and Conflict Resolution	0	0
18	Services	2	2
19	Social and Behavioral Science	5	6
20	Teacher Training	one	2
21	Veterinary		3
22	Other	TRUE	0
23	Total	255	
24	Number of Qualifications	20	
25	Minimum	0	
26	Maximum	69	
27			

17	Security and Conflict Resolution	0	0
18	Services	2	2
19	Social and Behavioral Science	5	6
20	Teacher Training	#NAME?	2
21	Veterinary		3
22	Other	TRUE	0
23	Total	#NAME?	
24	Number of Qualifications	20	
25	Minimum	#NAME?	
26	Maximum	#NAME?	
27			