I will build a categorical histogram and associated cumulative S-shape curve.

I will show counts of each category on the primary vertical axis and cumulative % values (from 0 to 100%) on the secondary vertical access.

As in many other examples, I have the first column random, and the second column is a paste special of it, paste values- which makes it fixed.

If we look at the data, we have the same categories: sometimes capitalized, sometimes all lowercase, and sometimes uppercase.

The POPER function capitalizes each word.

If we look at the PROPER column, some words have preceding blanks, and some may have a trailing blank

The TRIM function removes starting and trailing blanks.

If there are additional blanks between words, TRIM cannot remove them.

The UNIQUE function is a dynamic array form that lists the unique categories.

Under all dynamic array functions, we need enough empty space. Otherwise, we see the term SPILL.

Also, we see the formula in the first cell. Other cells are gray

I can order them by assigning 1, 2, 3.

I can then use dynamic array sort

When we use dynamic array sort, it is automatically updated

For example, let's change the order of poor and excellent

Let's undo it

Count if

different why they are different because

there is preceding a space or trailing

space over there what I can do I can use

a function trim to trim my original data

I'll go here I type trim original data

and I know I have this column I'll go

here i type trim and then I click on

this one if there is a preceding space

or trailing space this stream function

will drop it and this is a very safe

practice when we are dealing with big

data because when you prepare big data

you make a lot of mistakes in your typing

dream is one way one of the ways that we

can improve the quality of the data that

we have so I enter it and when I copied

down look this below average and this

this below average are different if here

effect

is this equal to this the answer is

false because the first one has a

proceeding space the second one doesn't

have so I go all the way down to trim

all my data now my data is trimmed okay

there is no proceeding space there is no

trailing space so I come here categories

now I am in a better position to use

that advanced property and identify

unique values again I will go to data

advanced I should move it to be so we

are talking about column B find the

unique values and copied to another

place it will ask us where do you want

me to copy it will say copied here I

click on column B then I say copy copy

in another place I click on unique

values and then I say ok and these data

are recorded over there now we sort this

data we go here we copy this data here

copy and then we sort this data go to

data and so did a to Z say continue on

this so then I will go here in the

frequency I type equal to count if when

I say count if I go and I select my

range and then because I am going to

copy down a 4 and this one count how

many above the average are there so I

click on it 9 then I copy down 5 average

three below average and one excellent

and two poor summation of these numbers

equal to sum and add them up

that is 20 is this frequency if you want

relative frequency each frequency should

be divided by summation of all numbers

and I lock the second one enter and then

I copy down that is relative frequency

in each category if you want cumulative

frequency I go here this one and then

here I say summation of this one f2 to

f3 but I make f2 absolute before enter

so that is point 7 then when I add it up

point eighty five point ninety point one

the last one is summation of all those

numbers so I was able to trim my data I

was able to create the categories I was

able to sort the categories and I was

able to compute the frequencies in

different categories now what I will do

I will go and select this data and I say

insert chart and I select this chart now

these numbers are there in the first

category I have nine observations in the

second five and so on the problem is

that here I have one two three four five

but what I want to have is below average

about the origin so I click on my balls

I right click select data edit

horizontal category and then in

horizontal category

I select these numbers okay and okay and

those number will appear here and the

categories are also appeared there one

other thing I do want to do is to

incorporate a cumulative frequency in

this graph so I click on one of the bars

select data add and when it asked for

add I will go to series value and I

click on these numbers okay and okay

then I go and select inside the chart

and say control one and then here for

this orange columns I say use the

secondary scale okay that is the

secondary scale

I have know the number of observations

is shown on the ax scale on the Left

which is numbers the cumulative

probability on the right then I go and

click on one of those columns change

series data type for the second one

instead of column I will use line okay

now this graph is over there demand my y

axis on the right hand side edit access

you don't need to go to one point two

because the maximum probability or

maximum relative frequency is 1 so

that's okay then it will go to 1 again I

missed format axis and I may say I don't

need you to go up with the steps of 0.1

let's go up with steps of say point 25

and that's it then I will have a graph

like this thank you very much