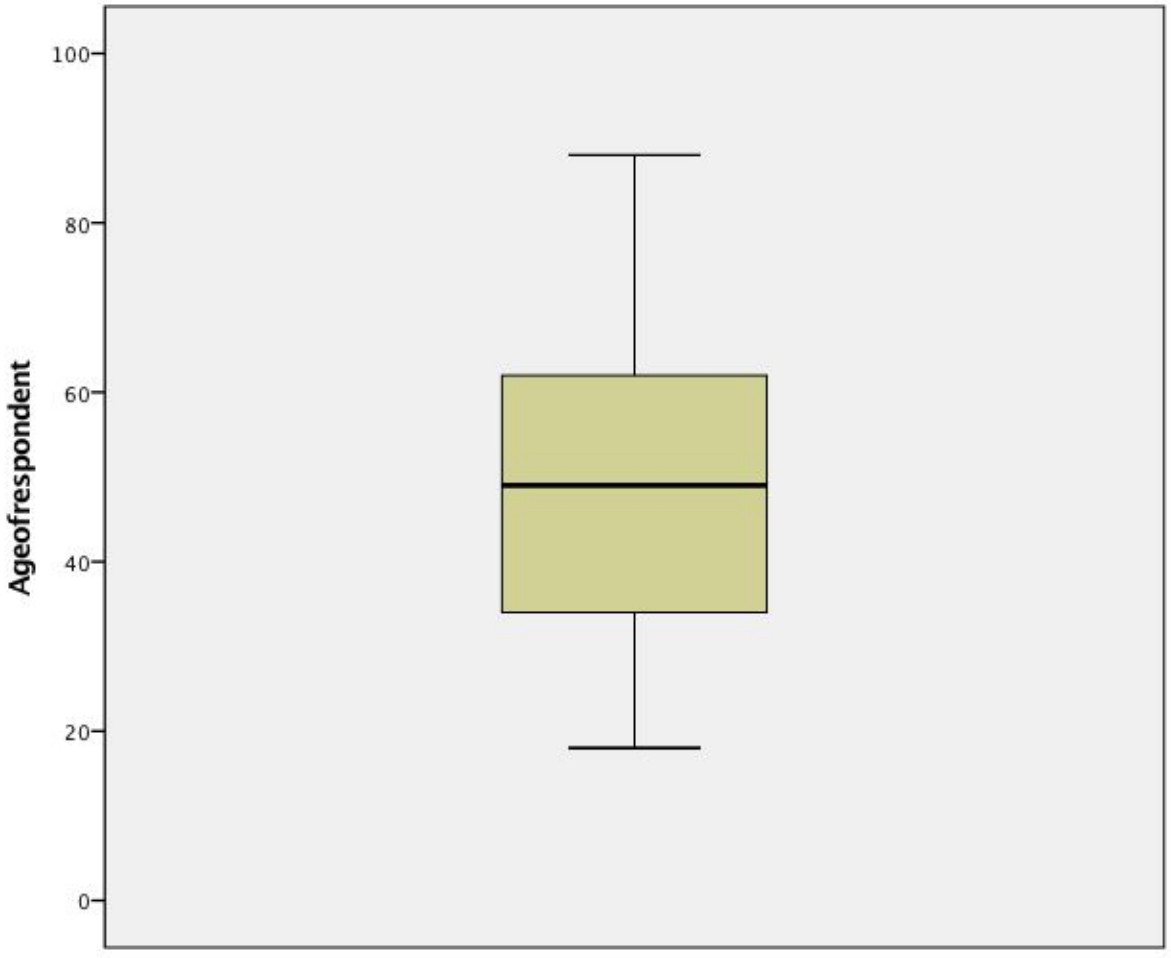
EUR2017 - HW4 - Appendix Worked Example

First I chose to ignore all the “not applicable” and “no answers”. I changed the data types of the number of bed partners variables to numeric.

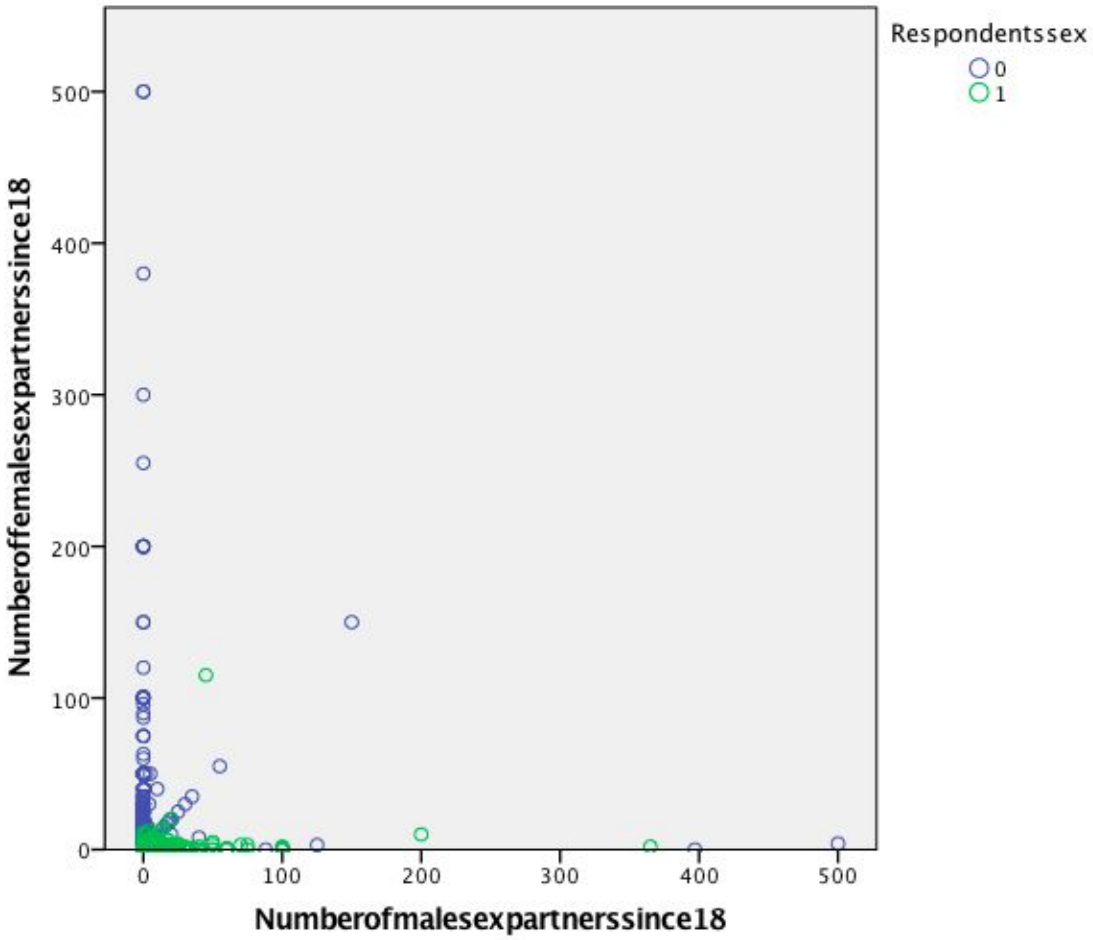
- There are 6 variables,
- 2870 respondents
- About 45% male, 55% female
- Data from one year: 2016

The age of respondents seems to be very balanced for such a large dataset:

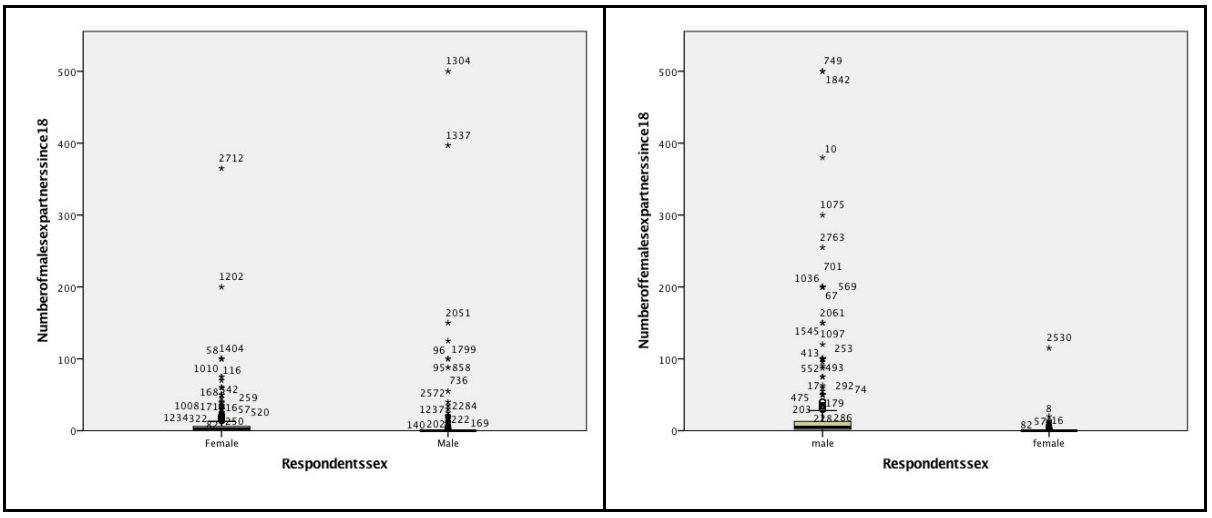


I created this graph which shows for males and females the number of male and female bed partners they said to have had. As you can see, most of females (green) report to have had a lot of male sex partners, and most males (blue) report to have had more female partners.

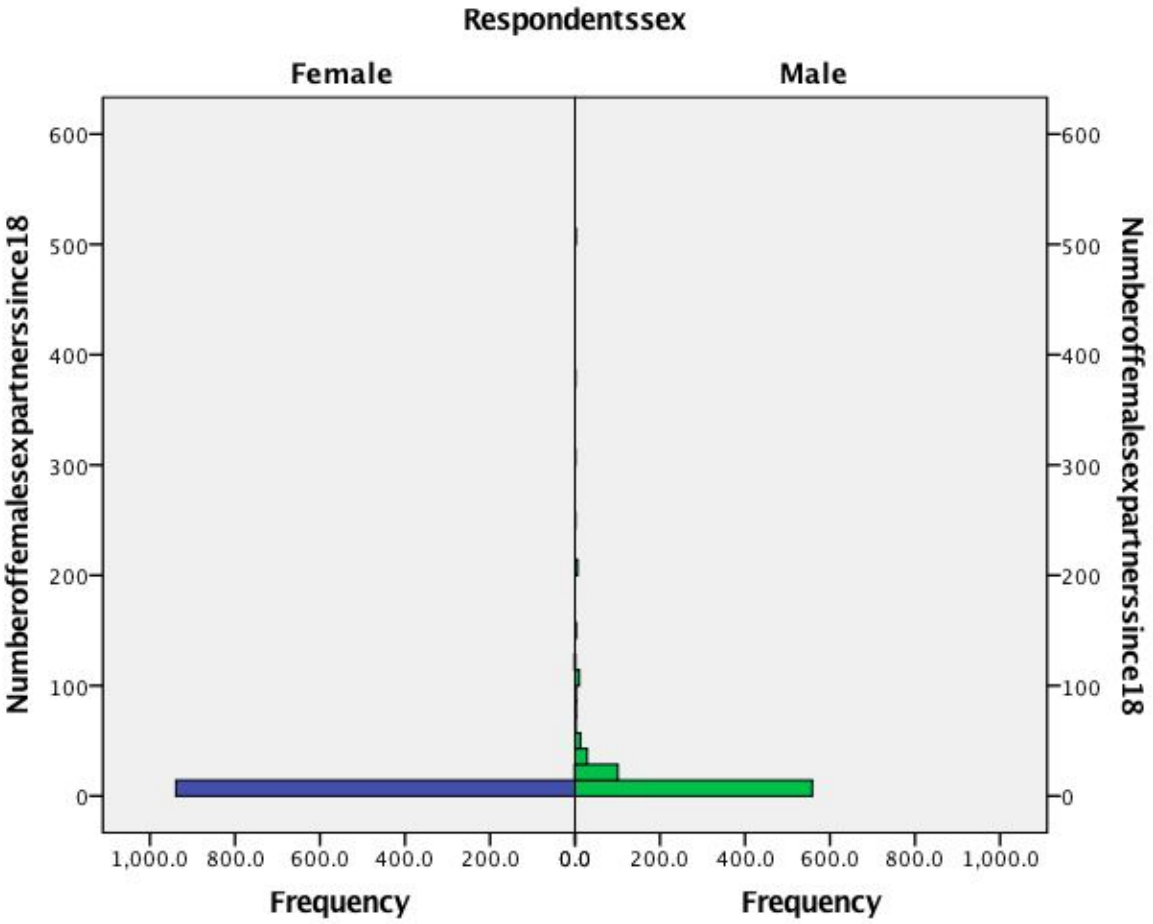
There are quite a few outliers, ranging from 100 to 500 sex partners.



A different way to visualise it is using box plots.



Or you can even get creative and use two-sided distributions:



Main take-away: most people are report to have had sex with handful of people. A couple of outliers said to have had sex with hundreds.

Insight: mean is probably not a good measure of centre in this case!

Now we want to compare the groups.
I go to Analyse > Compare Means > Means , and then put Gender as Independent Variable/ “Layer”

Output:

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Numberoffemalesexpartnerssince18 * Respondentssex	1675	58.4%	1195	41.6%	2870	100.0%
Numberofmalesexpartnerssince18 * Respondentssex	1667	58.1%	1203	41.9%	2870	100.0%

Report

Respondentssex		Numberoffemalesexpartnerssince18	Numberofmalesexpartnerssince18
male	Mean	15.54	2.68
	Median	5.00	.00
female	Mean	.49	6.65
	Median	.00	3.00
Total	Mean	7.08	4.82
	Median	.00	1.00

Apparently, based on the median, the average male REPORTS to have had sex with 5 (female) partners and the average female REPORTS to have had sex with 3 (male) partners.
Inter-sex bed partners don't show up.

Also interesting: the modes are the same for both sexes: 1. That's a lot different that the mean of 16 compared to a mean of 8 bed partners!

What I am now interested in is the balance. As this survey is said to be very representative, the number of bed partners on both sides should be somewhat in balance.
Let's observe the total counts.

	Male bedpartners	Female bedpartners
Sum	11862	8043

Close, but a bit unbalanced.

This makes me interested in the outliers.

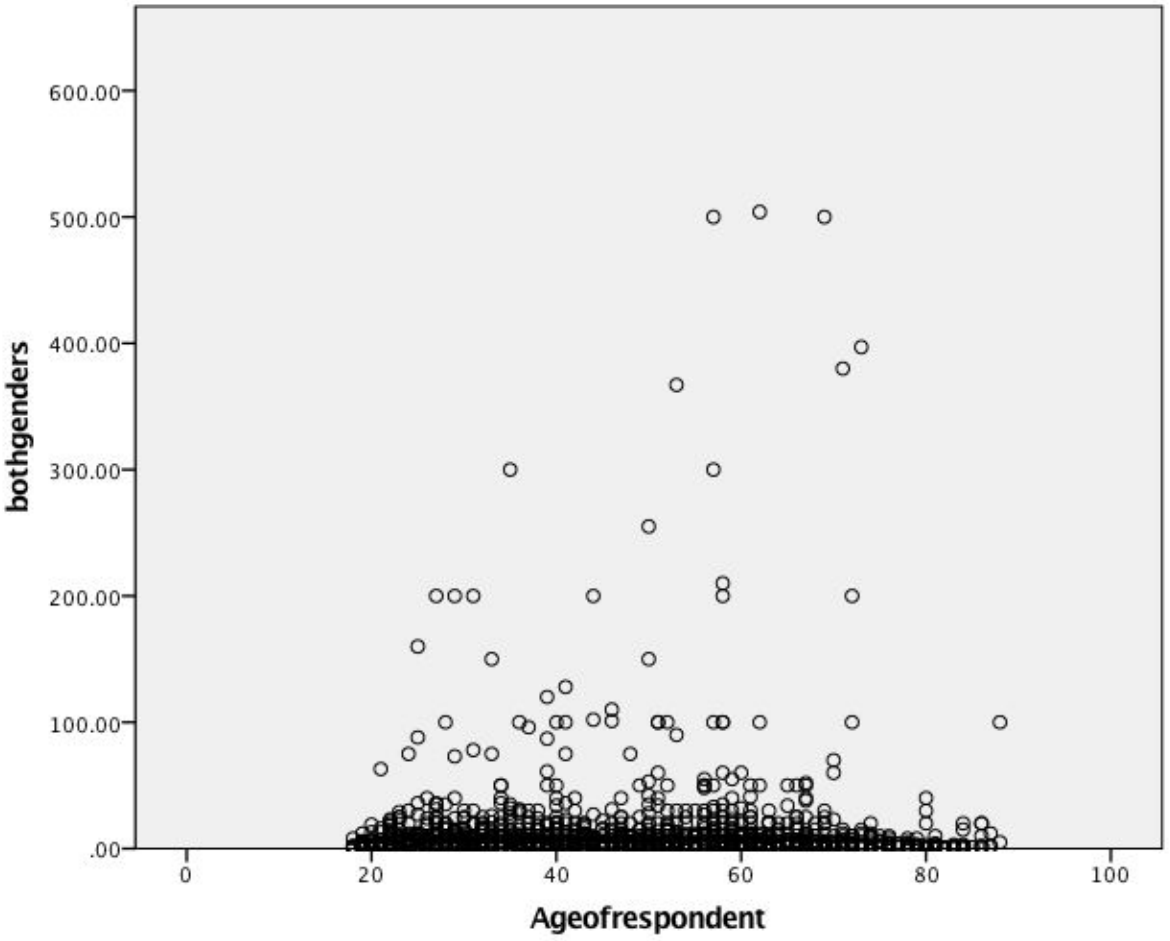
I do split file > Organize file based on groups , and select Gender as the grouping variable. This means that all our analyses will now be split by the grouping variable.

Now I do a frequency table, and look at the cumulative percentages:
Turns out that about 2% of men report to have slept with more than 100 females.

Only 0.2% of the women report to have slept with more than 100 males.

This could still be true, where the males would have slept with a lot of women, who have slept with only that man. But given that the sums are not equal, I suspect some men might be over reporting…

Who are these people that say to have slept with over 500 people?



Hopefully you have by now seen a lot of different stories you could write. Here are some examples:

Men report to have had twice as many bed partners

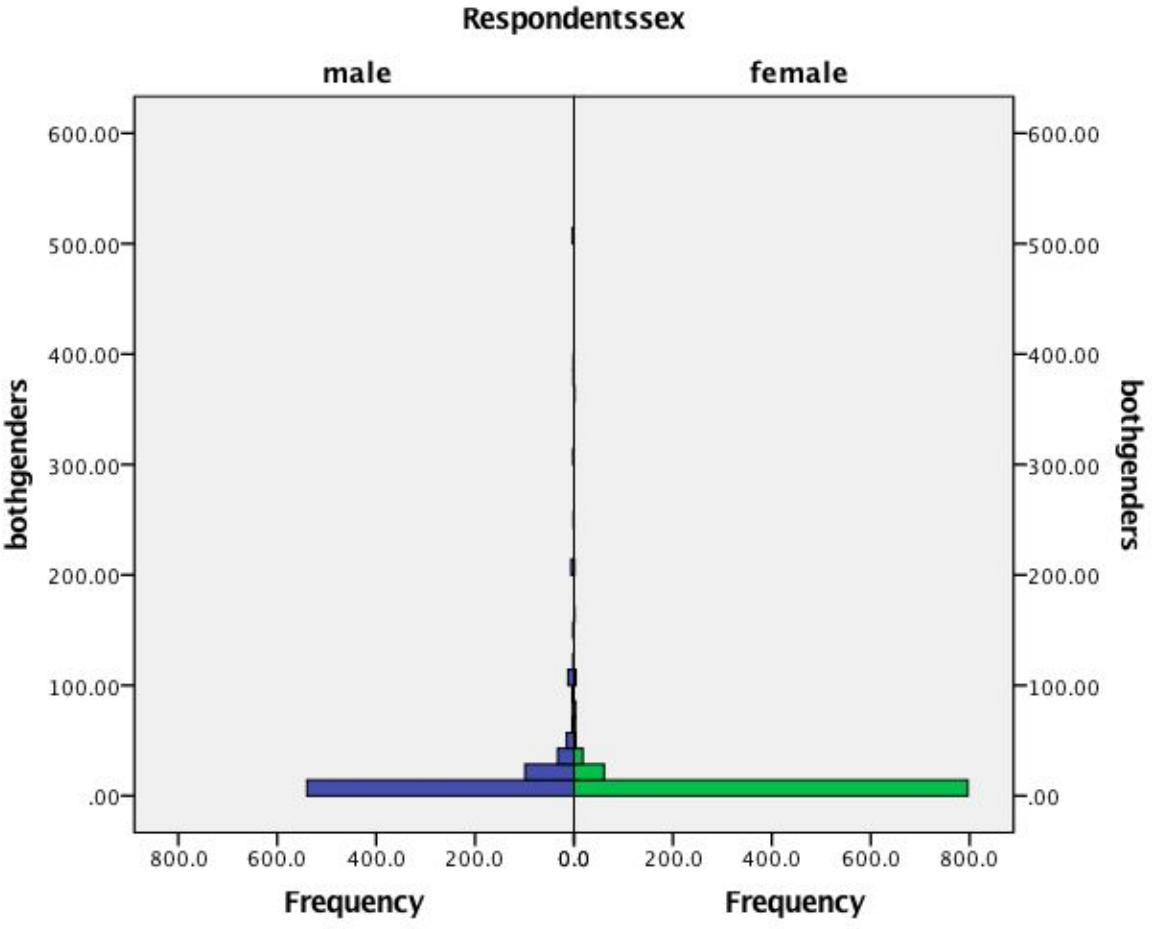
< this might not be such a good headline, as the mean is distorted by outliers.

Number of people you slept with: what is the mean(ing)?

Handful of men boast to have slept with hundreds.

Most people have had one to three bed partners.

In the US, sex can be counted on one hand



PS: Here I combined two variables. Using: Transform Variables > Compute Variable > bothgenders = numfemalepartners + nummalepartners

In a newspaper I would create a more detailed histogram of the lower part of the graph and then add the outliers all the way up to the top of page, showing how far apart the outliers are from the centre of the data.