Class 6: Analysis of bivariate data: frequency tables and graphical summaries

	Country						
	GB	Germany	France	Denmark	Sweden	Finland	Norway
Weighted Sample	1627	2038	1008	1023	1021	1002	583
Unweighted Sample	1627	2038	1008	1024	1016	1002	583
	%	%	%	%	%	%	%
At this point would you prefer that Britain stays in or leaves the European Union?							
Prefer that Britain stays in the EU	47	57	41	62	58	58	41
Prefer that Britain leaves the EU	44	17	30	17	21	19	27
Don't know	9	25	28	20	21	23	32

Objective

Up to now we have studied the characteristics of a single variable.

Often, various variables are measured at the same time.

As well as studying them individually, we are interested in whether there is a relationship between them

Preferred political party and wealth
Opinion on the death sentence and age
Opinion on Brexit and country

For qualitative and discrete data, data are often collected in a joint frequency table.

The two-way or joint frequency table

The following table is reconstructed from the results in a YouGov / Eurotrack survey in July 2018.

Who do you think has the upper hand in the Brexit negotiations?

		Country						
		GB	Germany	France	Denmark	Sweden	Finland	Norway
Opinion	Britain	163	489	322	215	255	250	122
	Rest of EU	1155	958	328	593	490	411	292
	Don't know	309	591	358	215	276	341	169

958 Germans said that the Rest of the EU has the upper hand.

341 Finns don't know.

The table with relative frequencies

		GB	Germany	France	Denmark	Sweden	Finland	Norway	
Opinion	Britain	163	489	322	215	255	250	122	
	Rest of EU	1155	958	328	593	490	411	292	
	Don't know	309	591	358	215	276	341	169	
									8302

Sum the data and …

divide by the sum.

		Country							
		GB	Germany	France	Denmark	Sweden	Finland	Norway	
Opinion	Britain	0.020	0.059	0.039	0.026	0.031	0.030	0.015	
	Rest of EU	0.139	0.115	0.040	0.071	0.059	0.050	0.035	
	Don't know	0.037	0.071	0.043	0.026	0.033	0.041	0.020	
									1.000

958/8302 = 11.5% of the sample are Germans who think the EU has the upper hand.

341/8302 = 4.1% of the people surveyed are Finns who don't know.

Marginal frequencies

		Country							
		GB	Germany	France	Denmark	Sweden	Finland	Norway	Total
Opinion	Britain	163	489	322	215	255	250	122	1816
	Rest of EU	1155	958	328	593	490	411	292	4227
	Don't know	309	591	358	215	276	341	169	2259
	Total	1627	2038	1008	1023	1021	1002	583	8302

1627 = 163 + 1155 + 309 people in the survey are from GB.

		Country							
		GB	Germany	France	Denmark	Sweden	Finland	Norway	Total
Opinion	Britain	0.020	0.059	0.039	0.026	0.031	0.030	0.015	0.219
	Rest of EU	0.139	0.115	0.040	0.071	0.059	0.050	0.035	0.509
	Don't know	0.037	0.071	0.043	0.026	0.033	0.041	0.020	0.272
	Total	0.196	0.245	0.121	0.123	0.123	0.121	0.070	1.000

Just over half of the people surveyed think the rest of the EU has the upper hand.

Conditional frequencies

		Country						
		GB	Germany	France	Denmark	Sweden	Finland	Norway
Opinion	Britain	163	489	322	215	255	250	122
	Rest of EU	1155	958	328	593	490	411	292
	Don't know	309	591	358	215	276	341	169
	Total	1627	2038	1008	1023	1021	1002	583

163 out of 1627 GB residents think Britain has the upper hand in the Brexit negotiations.

		Country						
		GB	Germany	France	Denmark	Sweden	Finland	Norway
Opinion	Britain	0.100	0.240	0.319	0.210	0.250	0.250	0.209
	Rest of EU	0.710	0.470	0.325	0.580	0.480	0.410	0.501
	Don't know	0.190	0.290	0.355	0.210	0.270	0.340	0.290
	Total	1.000	1.000	1.000	1.000	1.000	1.000	1.000

163/1627 ≈ 10% of the GB residents in the survey think that Britain has the upper hand.

Conditional frequencies

		Country							
		GB	Germany	France	Denmark	Sweden	Finland	Norway	Total
Opinion	Britain	0.020	0.059	0.039	0.026	0.031	0.030	0.015	0.219
	Rest of EU	0.139	0.115	0.040	0.071	0.059	0.050	0.035	0.509
	Don't know	0.037	0.071	0.043	0.026	0.033	0.041	0.020	0.272

21.9% of the people surveyed think Britain has the upper hand.

2% of the people surveyed are from GB and think Britain has the upper hand.

		Country							
		GB	Germany	France	Denmark	Sweden	Finland	Norway	Total
Opinion	Britain	0.090	0.269	0.177	0.118	0.140	0.138	0.067	1.000
	Rest of EU	0.273	0.227	0.078	0.140	0.116	0.097	0.069	1.000
	Don't know	0.137	0.262	0.158	0.095	0.122	0.151	0.075	1.000

0.020/0.219 = 9% of the people who think Britain has the upper hand in negotiations come from GB. Just over ¼ of the people who think Britain has the upper hand are Germans.

Checking for relations between variables

Do European countries have different opinions about who is dominating Brexit negotiations?

Overall, 21.9% of the people surveyed think Britain is doing best and 50.9% think the rest of the EU.

		Total
Opinion	Britain	0.219
	Rest of EU	0.509
	Don't know	0.272
	Total	1.000

If opinion and country were related, we would expect to see approximately these values for all countries. However, it seems GB residents are more pessimistic than people from other EU countries and the French think Britain is doing better than other countries.

		Country						
		GB	Germany	France	Denmark	Sweden	Finland	Norway
Opinion	Britain	0.100	0.240	0.319	0.210	0.250	0.250	0.209
	Rest of EU	0.710	0.470	0.325	0.580	0.480	0.410	0.501
	Don't know	0.190	0.290	0.355	0.210	0.270	0.340	0.290
	Total	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Look out!

Nearly all tables that we see in the news, government reports, … are conditional frequency tables.

YouGov / Eurotrack Survey Results

Sample Size: 1627 GB Adults / 2038 German Adults / 1008 French Adults / 1024 Danish Adults / 1016 Swedish Adults / 1002 Finnish Adults / 583 Norwegian Adults
Fieldwork: 19th - 26th July

	Country						
	GB	Germany	France	Denmark	Sweden	Finland	Norway
Weighted Sample	1627	2038	1008	1023	1021	1002	583
Unweighted Sample	1627	2038	1008	1024	1016	1002	583
	%	%	%	%	%	%	%
Who do you think has the upper hand in the Brexit negotiations?							
Britain	10	24	32	21	25	25	21
Rest of EU	71	47	33	58	48	41	50
Don't know	19	29	36	21	27	34	29

How can you tell if a table contains joint frequencies or conditional frequencies?

Exercise

A statistical institute has carried out a survey to predict the voting habits of first time voters, currently of ages 18 to 20 in the next elections in the UK. The number of people sampled was 3000. It is wished to study the relationship between the intention to vote and age. Let X = Political party and Y = age of the respondent:

	18	19	20
Conservative	450	300	210
Labour	250	270	330
Liberal	145	170	200
Nationalist groups	95	150	180
Independents	50	115	85

Which one of the following options is correct?

a) 33.5% of the people sampled are under 20 years old.
b) 5.67% of the people sampled are 19 year old Liberal voters.
c) 10% of the people sampled are Independents.
d) 10% of 19 year olds are Conservative voters.

Exercise

Following from the previous question, signal the correct answer.

a) 31.76% of the people who intend to vote Labour are 20 years old.
b) 29.41% of the prospective Labour voters are 19 years old.
c) 29.41% of the prospective Labour voters are 18 years old.
d) 38.82% of the people who intend to vote Labour are 19 years old.

	18	19	20
Conservative	450	300	210
Labour	250	270	330
Liberal	145	170	200
Nationalist groups	95	150	180
Independents	50	115	85

Exercise

The following table is taken from the same YouGov survey as the example studied earlier and reflects the opinions of the respondents on whether or not they think their own country should leave the EU.

	Country						
	GB	Germany	France	Denmark	Sweden	Finland	Norway
Weighted Sample	1627	2038	1008	1023	1021	1002	583
Unweighted Sample	1627	2038	1008	1024	1016	1002	583
	%	%	%	%	%	%	%
If there was a referendum on COUNTRY'S membership of the European Union, how would you vote?							
I would vote to remain a member of the European Union	47	55	49	60	56	53	20
I would vote to leave the European Union	41	23	26	26	28	28	62
I would not vote	5	7	9	2	2	7	5
Don't know	7	13	15	11	13	11	13
Refused	1	2	1	1	1	0	0

Which of the following is correct?
(a) 62% of the people surveyed are Norwegian and want Norway to leave the EU.
(b) 62% of the people in the survey who want their country to leave the EU are Norwegians.
(c) 62% of the Norwegians in the survey want Norway to leave the EU.
(d) None of the previous answers.

Exercise

The table below is taken from the CIS barometer of June 2018.

Pregunta 30

Le voy a presentar ahora algunas fórmulas alternativas de organización territorial del Estado en España. Dígame, por favor, con cuál está Ud. más de acuerdo

	TOTAL	Edad de la persona entrevistada					
		De 18 a 24 años	De 25 a 34 años	De 35 a 44 años	De 45 a 54 años	De 55 a 64 años	65 y más años
Un Estado con un único Gobierno central sin autonomías	21,1	9,3	14,7	19,8	23,1	23,0	26,6
Un Estado en el que las comunidades autónomas tengan menor autonomía que en la actualidad	9,8	7,7	8,8	10,6	12,3	9,9	8,2
Un Estado con comunidades autónomas como en la actualidad	34,0	42,3	36,0	34,3	32,7	33,4	31,4
Un Estado en el que las comunidades autónomas tengan mayor autonomía que en la actualidad	13,1	18,0	19,5	14,7	11,7	11,7	9,0
Un Estado en el que se reconociese a las comunidades autónomas la posibilidad de convertirse en Estados independientes	9,2	9,8	9,4	10,0	12,9	9,4	5,5
N.S.	10,5	11,9	10,3	7,7	5,6	9,1	16,9
N.C.	2,3	1,0	1,2	3,0	1,7	3,4	2,4
(N)	(2.487)	(194)	(339)	(470)	(480)	(383)	(621)

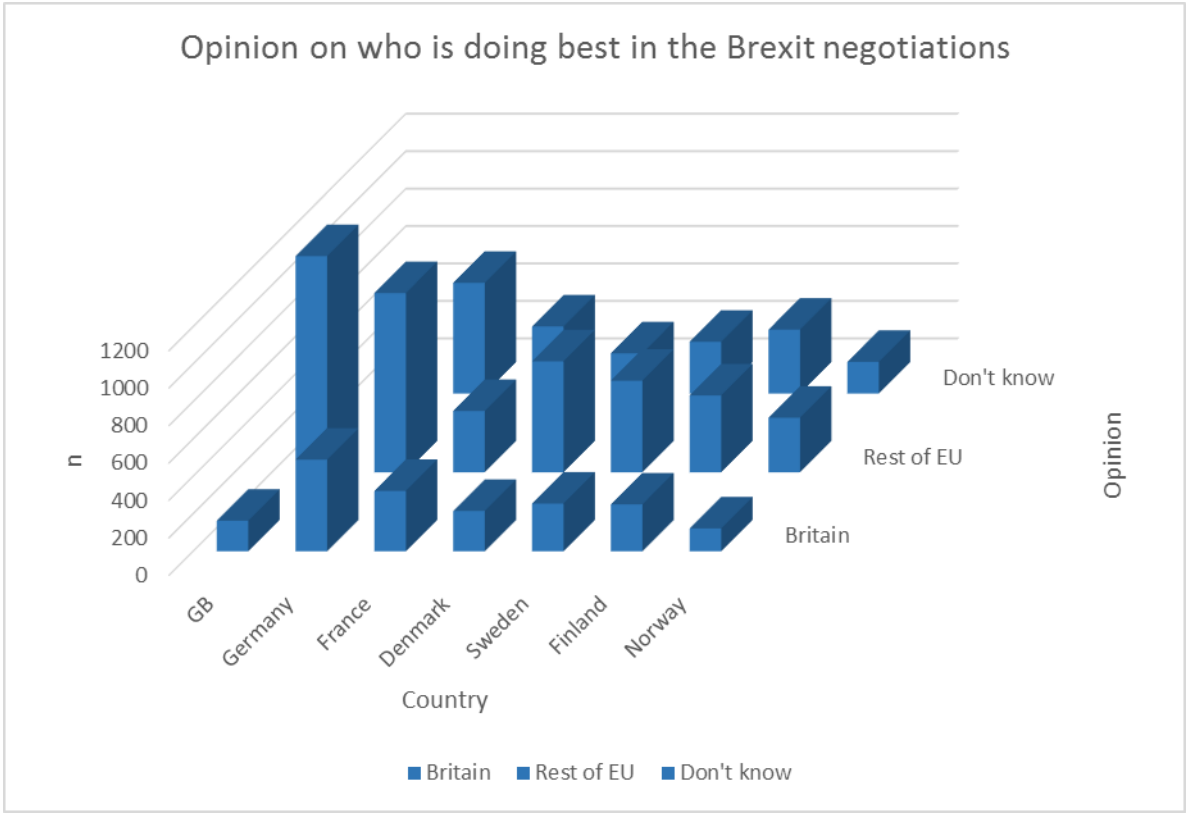
a) What proportion of the people surveyed think that Spain should just be run by a central government without autonomous communities?
b) What proportion of the under 25s think that Spain should just be run by a central government.
c) Out of the people who think that Spain should be run by a central government, what proportion of them are under 25?

Graphical summaries

3d bar charts

could be used to display the joint frequencies of a bivariate distribution of two qualitative variables.

What are the advantages and disadvantages?

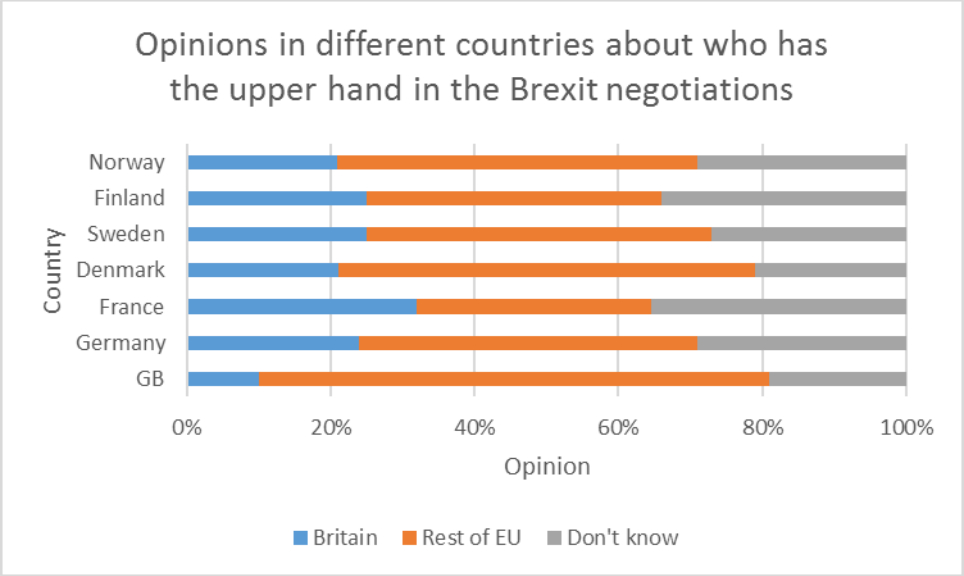


Graphical summaries

Stacked bar charts

can be used to display the conditional distributions for qualitative variables.

Opinions in GB and France look quite different!

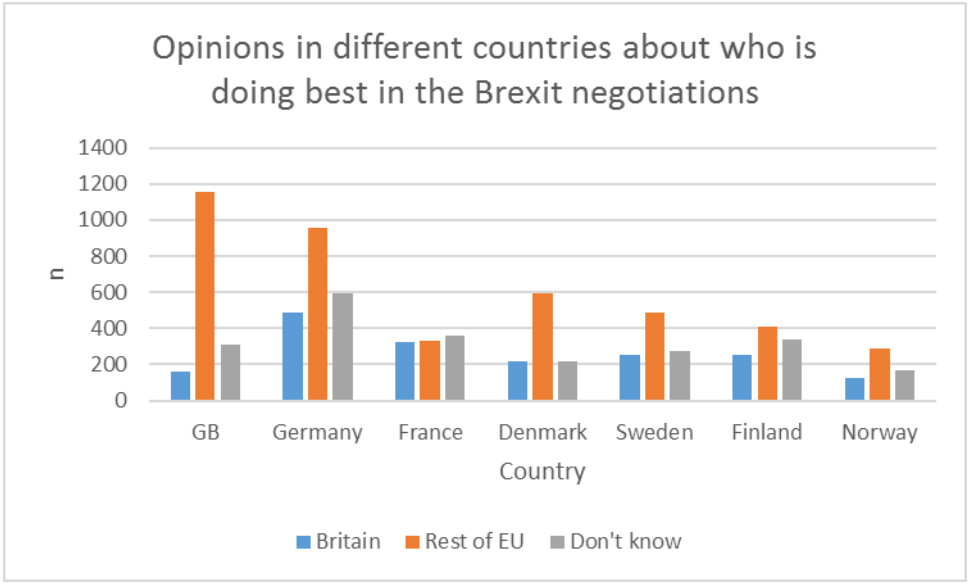


Graphical summaries

Multiple bar charts

can be used to display absolute frequencies as well as differences in opinions between groups.

The shape of the distributions in GB and France is different. Many more people from GB have been surveyed.

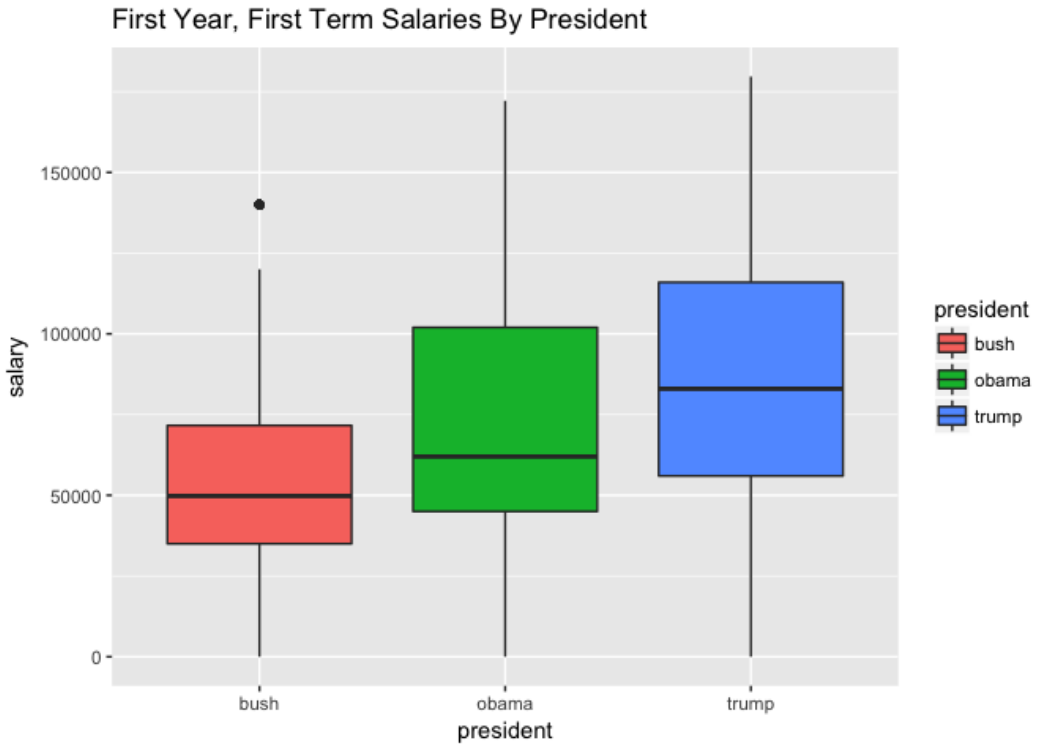


Graphical summaries

Multiple box plots

can be used to show the differences in the distribution of a continuous variable between various groups.

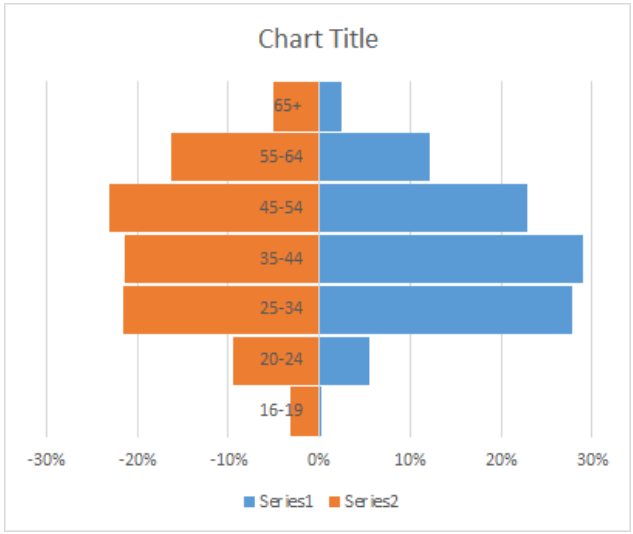
The graphs show the distributions of salaries of all White House employees under different presidents.



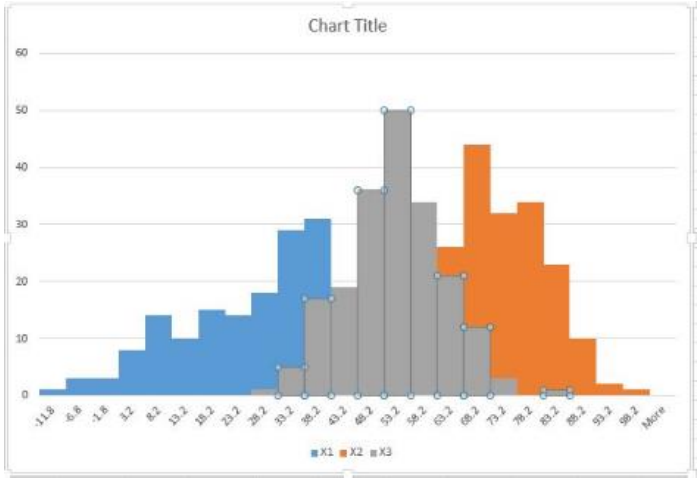
Graphical summaries

Back to back or overlaid histograms

can be used to compare the distribution of a continuous variable between in two …



or more than two groups.

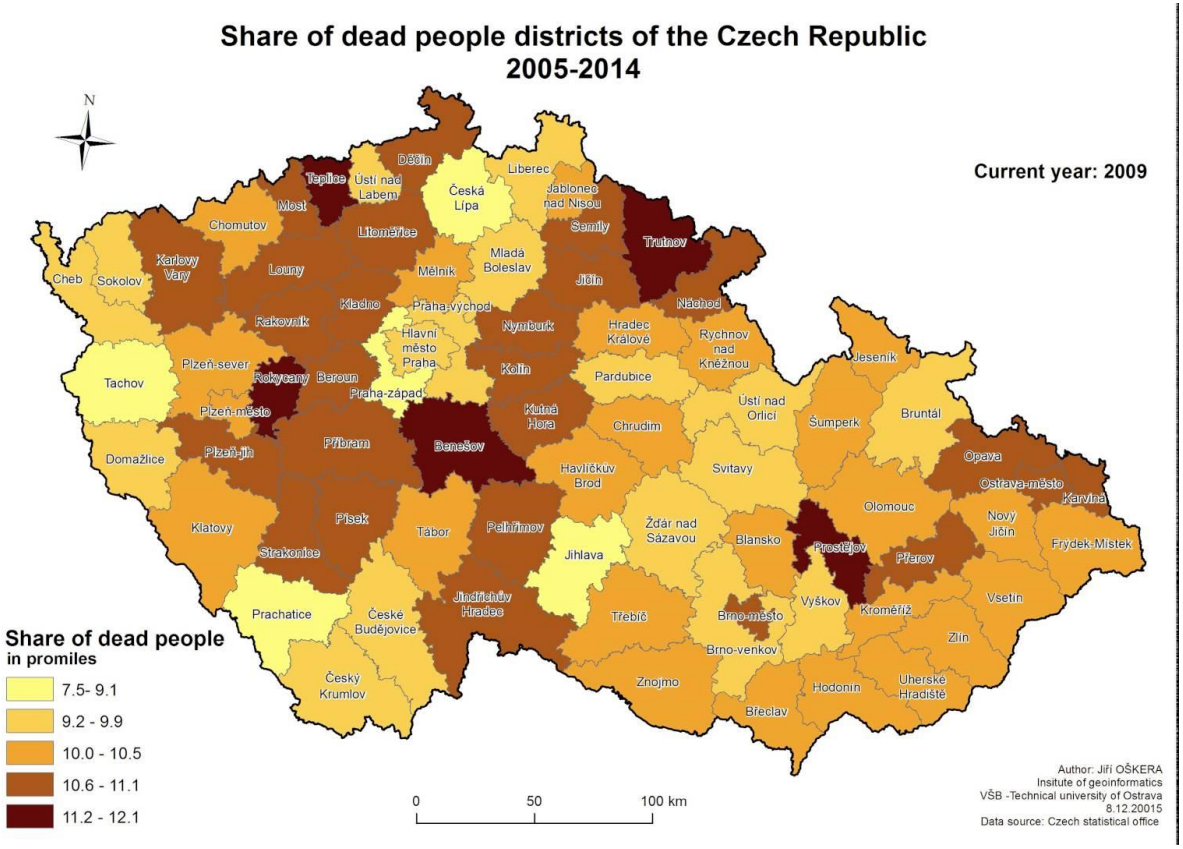


Graphical summaries

Cartograms

can be used to show the differences in the distribution of a continuous variable over different locations.

The graph shows the number of deaths per 1000 inhabitants in different regions. Darker colours indicate higher death rates.



Graphical summaries

Scatterplots

can be used to show the relationship between two quantitative variables.

It looks like there is an almost linear relation between seats and population (in 2016).