



Class 2: Analysis of univariate data: graphics for qualitative data



Recommended reading:

[This page](#) in Wikipedia explains how to mislead with different graphics.



Example

SAMPLE: UC3M students

VARIABLE: Preferred Political Party

OBJECTIVE: Classification and representation of information

Partido político preferido		
PP	PP	PP
PP	PSOE	PSOE
PSOE	PP	PP
Podemos	PP	Cs
Cs	Cs	Podemos
Other	Cs	Cs
PSOE	Podemos	PSOE
PP	Podemos	PP
Cs	Other	PP
PP	Other	PSOE
Other	PSOE	Cs
Podemos	PSOE	Other
Podemos	PP	PP
PP	Cs	Cs
Cs	Podemos	Other
PSOE	PP	PSOE
PSOE	PP	PP
Podemos	Cs	PP
Other	Podemos	PSOE
Podemos	Other	PSOE
Podemos	PP	Podemos



The frequency table

Political Party	Absolute frequency (n)	Relative frequency (f)
PP	19	0.30
PSOE	13	0.21
Podemos	12	0.19
Cs	11	0.17
Other	8	0.13
Total	63	1.00

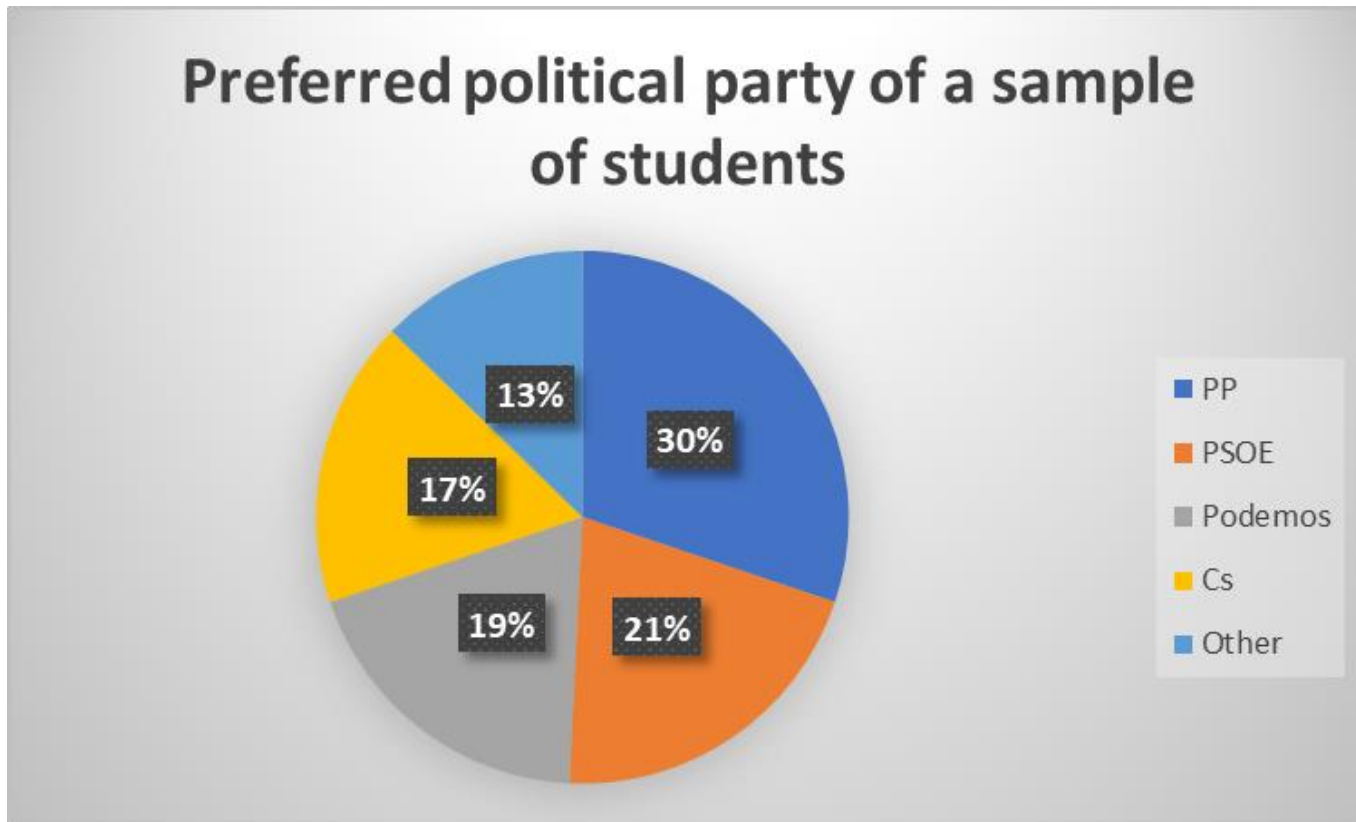
12 Podemos voters in the sample

= 19/63

= 19 + 13 + ... + 8



The pie chart



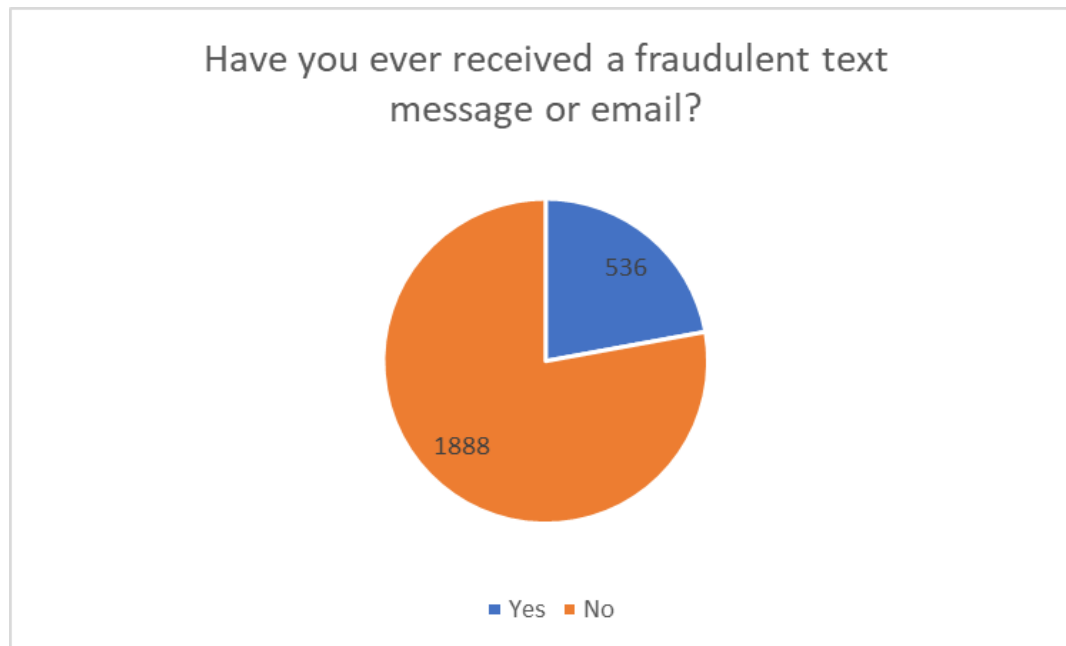


When should we use a pie chart?



Dichotomous or polytomous data with few categories

Q18 In the last 12 months have you received a fraudulent text message or email?





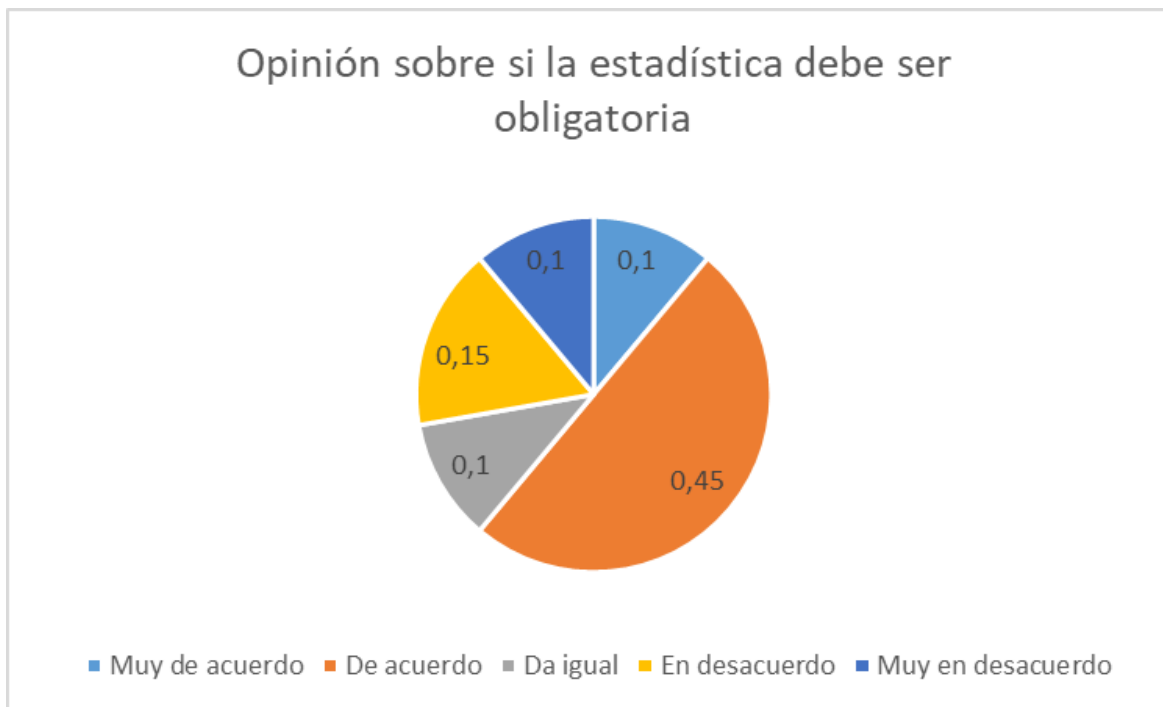
When shouldn't we use a pie chart?



Quantitative data: the distance between values isn't represented



Ordinal data: the sense of order is lost



The percentages are correct but ...



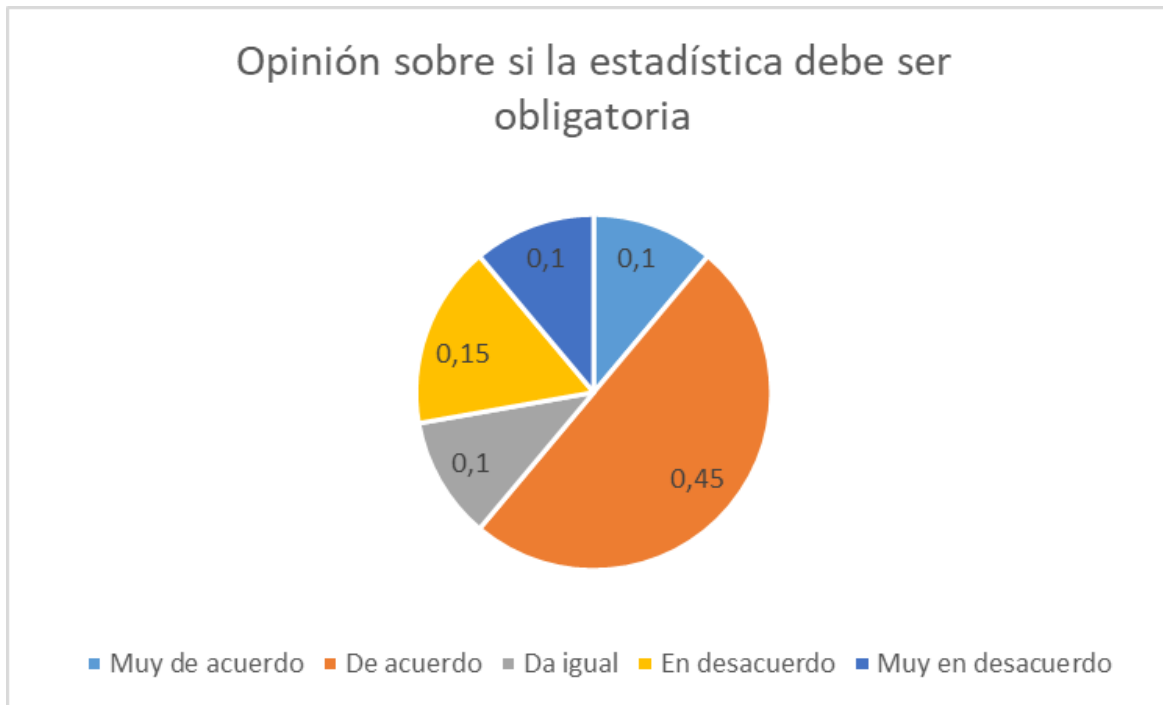
When shouldn't we use a pie chart?



Quantitative data: the distance between values isn't represented



Ordinal data: the sense of order is lost



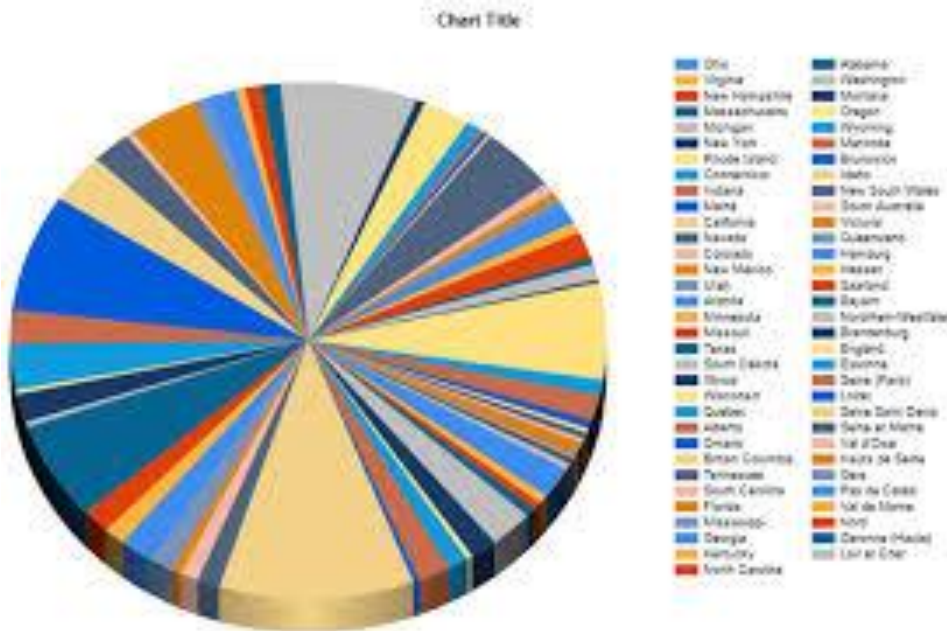
Very much against next to very much in favor?



When we shouldn't use a pie chart



Nominal data with lots of categories



Makes me feel like I've done too many drugs!

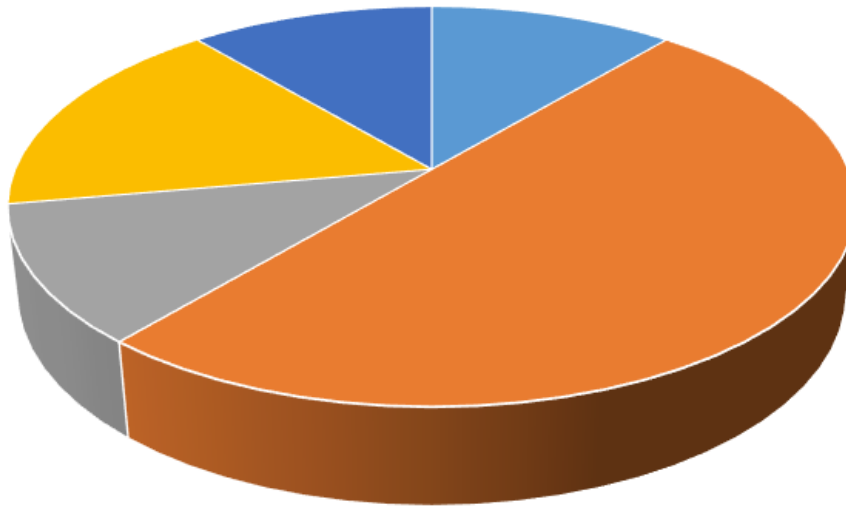


Stupid pie charts



Diagrams in 3d

Opinión sobre si la estadística debe ser obligatoria

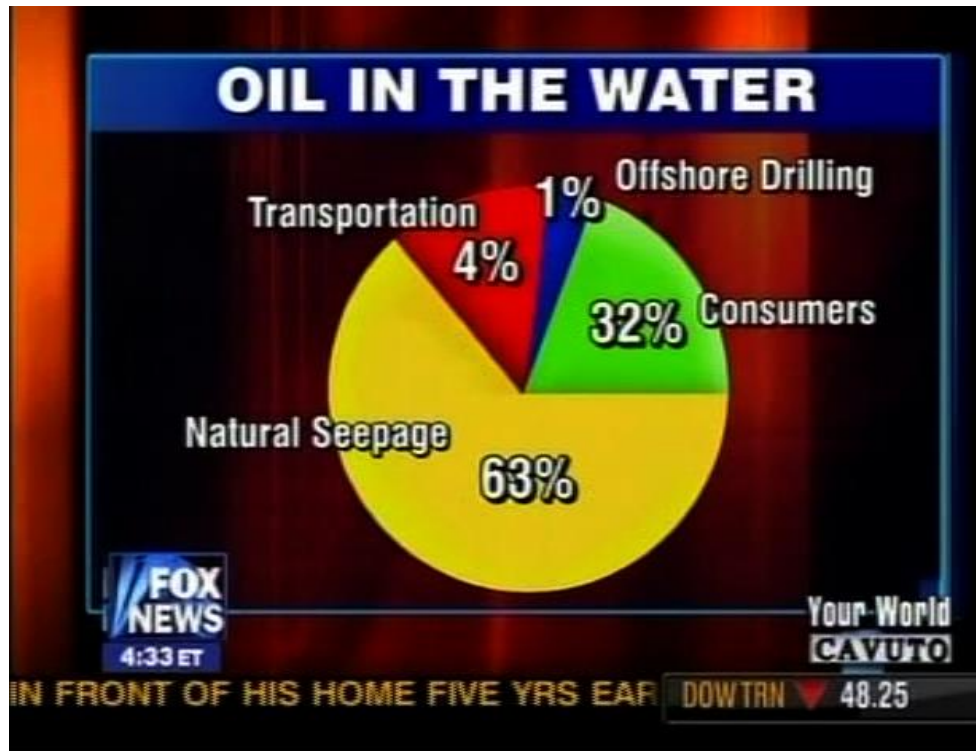


■ Muy de acuerdo ■ De acuerdo ■ Da igual ■ En desacuerdo ■ Muy en desacuerdo

It looks like there are more “don’t knows” than “against”. In reality the proportions are the same.



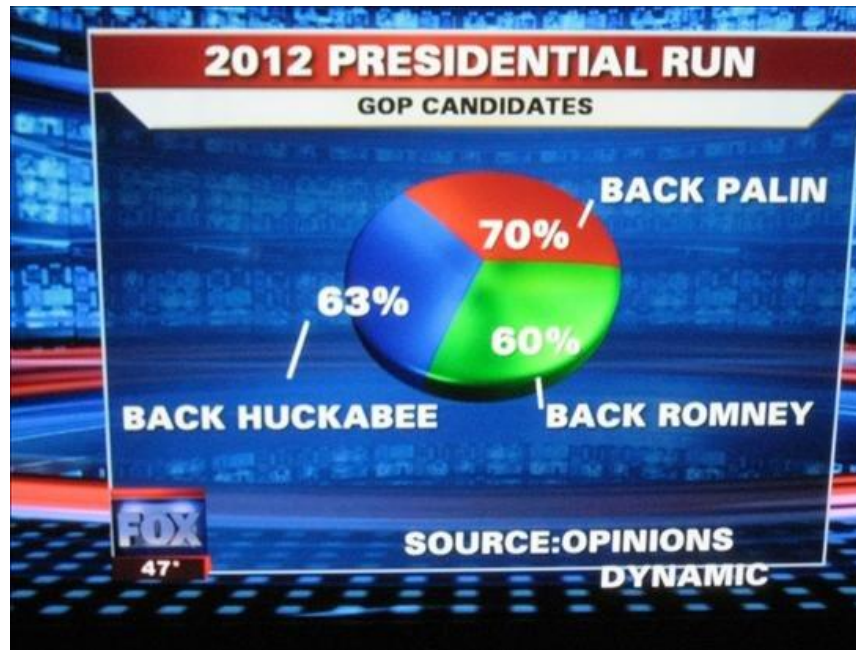
Stupid pie charts



Fox News: Trump's preferred news channel!



Stupid pie charts



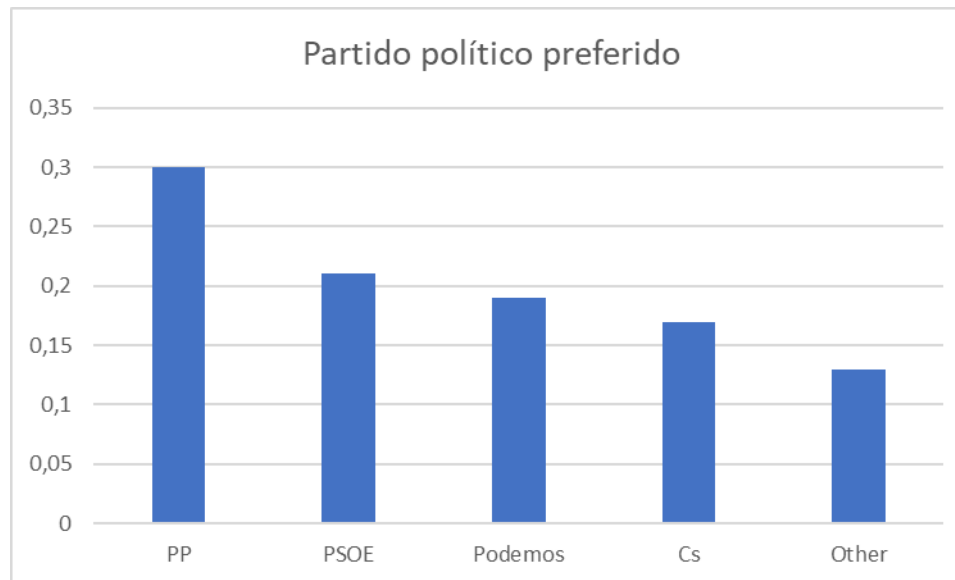
Fox News again!



The bar chart



Use with qualitative (or discrete) data.
With ordinal data it is important to preserve the sense of order.





Lying with bar charts

Venezolana de Televisión

VENEZUELA *¡de verdad!*

Lunes, 15 de Abril 2013

Buscar en el Sitio

SEÑAL EN VIVO RADIOWEB VTV

INICIO NOTICIAS PROGRAMAS MULTIMEDIA OPINIÓN EL CANAL ELECCIONES HUGO CHÁVEZ

VOTO 2013 ELECCIÓN PRESIDENCIAL 2013
El pueblo decide

PORCENTAJES

Candidate	Percentage
Nicolás Maduro Moros	50,66%
Henrique Capriles Radonski	49,07%

Ganó el hijo de Chávez: Nicolás Maduro es el Presidente Electo de Venezuela (+Video)

ELECCIONES

La tendencia es irreversible, señaló Tibisay Lucena, presidenta del ente comicial / Maduro obtuvo 50,66% y Capriles 49,07%

[Leer más](#)

Nicolás Maduro Moros

Henrique Capriles Radonski

Ganó el hijo de Chávez: Nicolás Maduro es el Presidente Electo de Venezuela (+Video)

Nicolás Maduro rechazó "Falso" que le propuso Capriles antes de conocerse resultados oficiales

Hackeadas cuentas Twitter del Presidente Nicolás Maduro y del País

Nicolás Maduro: Hoy es un día de historia, vamos con Cristo Redentor (+Fotos)

Televisión Venezolana



Pictograms



The **area** of the graph is proportional to the frequency.



Lying with pictograms



People see areas
and not just heights!



Word clouds



Word clouds are a type of pictogram used to illustrate the frequency of certain words in a text.

What are the [advantages and disadvantages of Word clouds?](#)



Exercise

<u>Type of crime</u>	N°
Petty theft	5996
Assault	5059
Damages	3866
Narcotics	3134
Robbery	2289
Aggression	1641
Others	1802

Crimes committed in Chicago between 5/11/2009 and 5/12/2009.

Draw an appropriate graph of these data



Exercise

Final grades in a Statistics course:

<u>CLASS</u>	Nº Students
No presentado	17
Suspenso	16
Aprobado	29
Notable	24
Sobresaliente	2

Is the variable nominal or ordinal?

What percentage of the students didn't pass?

What would be a good graphic for these data?



Exercise

The 40 students in a statistics class rate their lecturer from 1 (boring) to 5 (fantastic). The following table shows the partial results of the survey.

Evaluation	Absolute Frequency	Relative Frequency
1		0,05
2		
3	5	
4	9	
5	19	
TOTAL		

Complete the table.



Exercise

The results reveal that Britons would prefer David Cameron to be in charge than Theresa May, with 28% plumping for the former and 22% the latter in a head to head. Nevertheless, 41% of Brits answered “neither” and a further 9% “don’t know”.



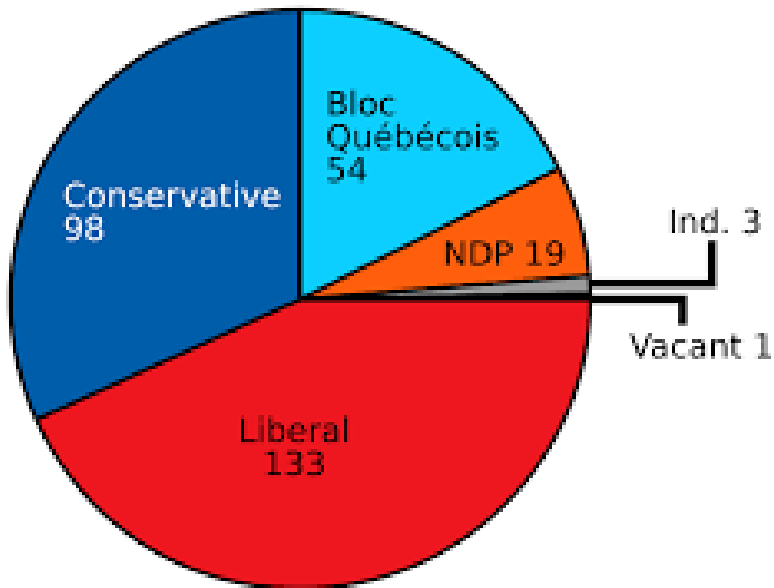
Theresa May vs
David Cameron

Briefly comment on the above graphic.



Exercise

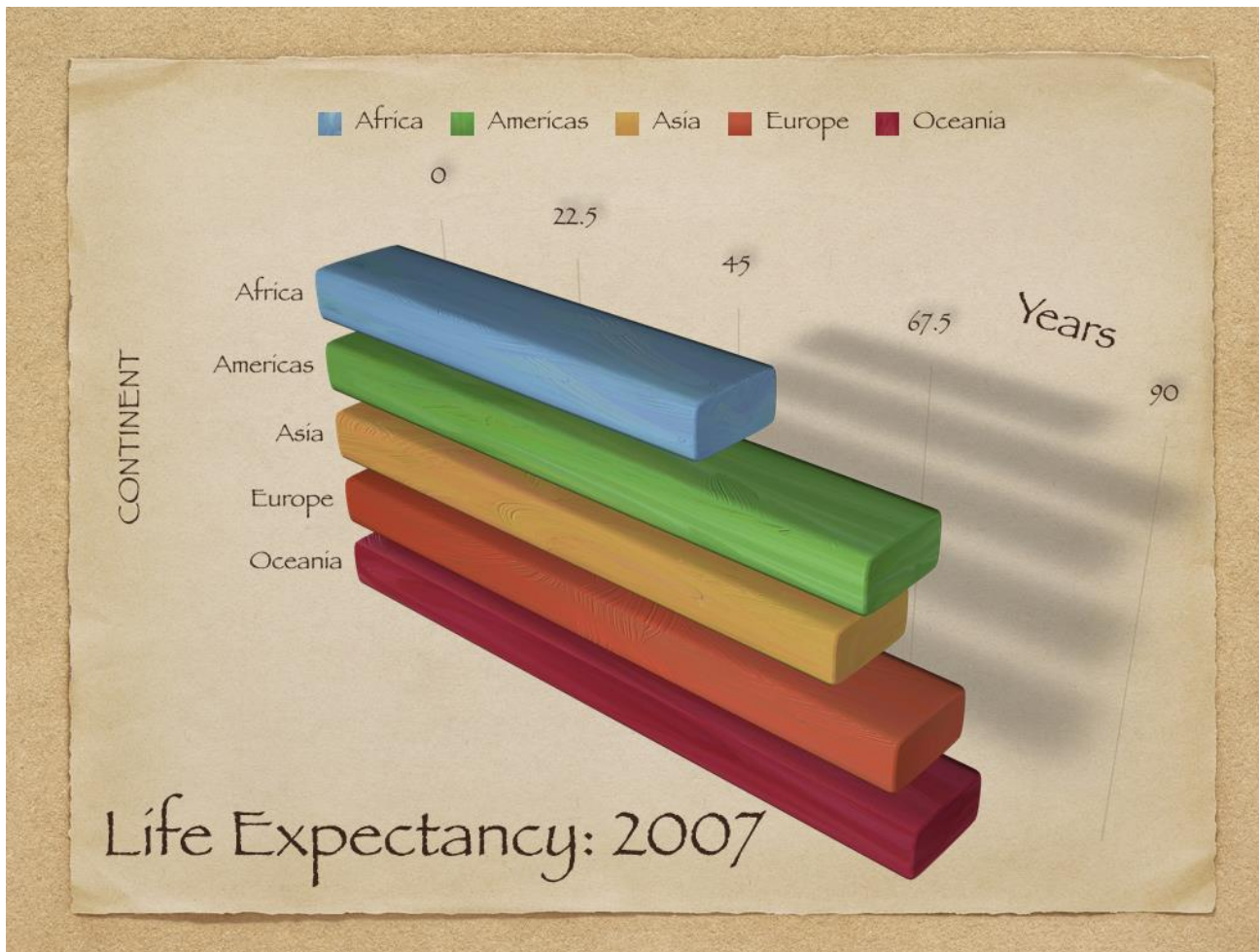
Composition of 38th Parliament of Canada as of May 19, 2005



The pie chart reflects the formation of the Canadian Parliament in 2005. Comment briefly on the graph and any possible alternatives.



Exercise



The graphic indicates life expectancy in 2007 in the different continents. Comment briefly on the suitability of the graph and any possible alternatives.