Chapter 2

2.1 Nominal: Occupation, undergraduate major. Ordinal: Rating of university professor, Taste test ratings. Interval: age, income

2.2 a Interval

b Interval

c Nominal

d Ordinal

2.3 a Interval

b Nominal

c Ordinal

d Interval

e Interval

2.4 a Nominal

b Interval

c Nominal

d Interval

e Ordinal

2.5 a Interval

b Interval

c Nominal

d Interval

e Nominal

2.6 a Interval

b Interval

c Nominal

d Ordinal

e Interval

2.7 a Interval

b Nominal

c. Nominal

d Interval

e Interval

f Ordinal

2.8 a Interval

b Ordinal

c Nominal

d Ordinal

2.9 a Interval

b Nominal

c Nominal

2.10 a Ordinal

b Ordinal

c Ordinal

2.11 a Nominal

b Interval

c Ordinal

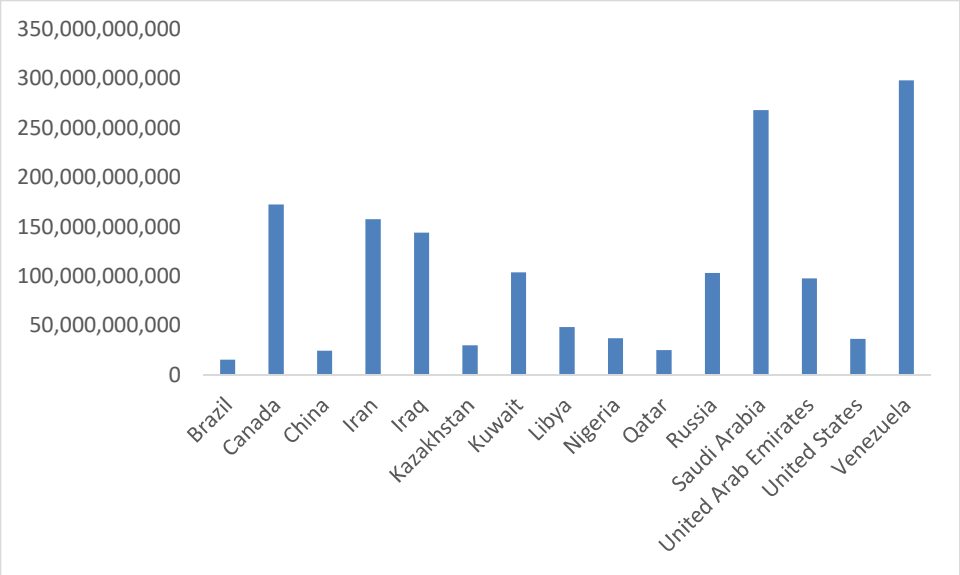
2.12a Nominal

b Interval

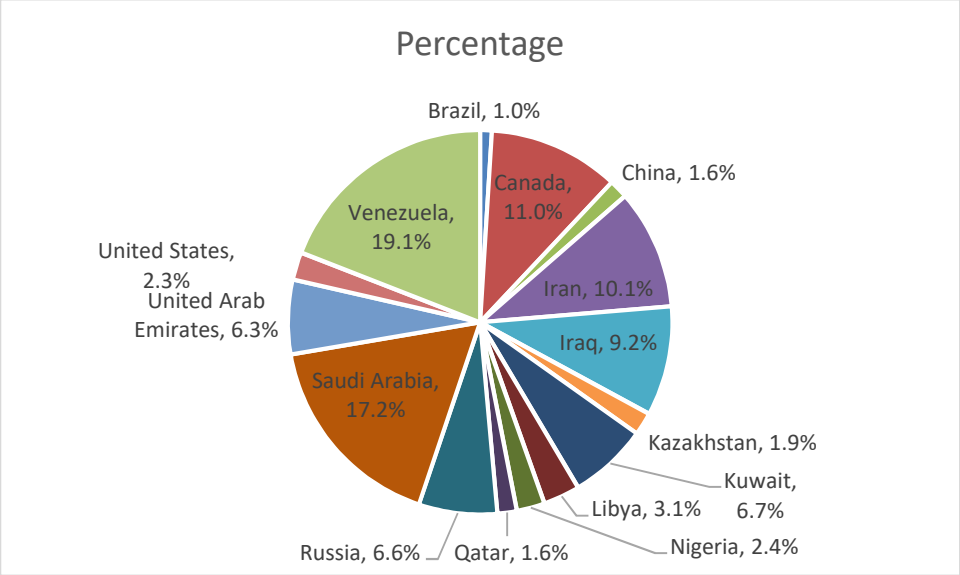
c Interval

d Interval

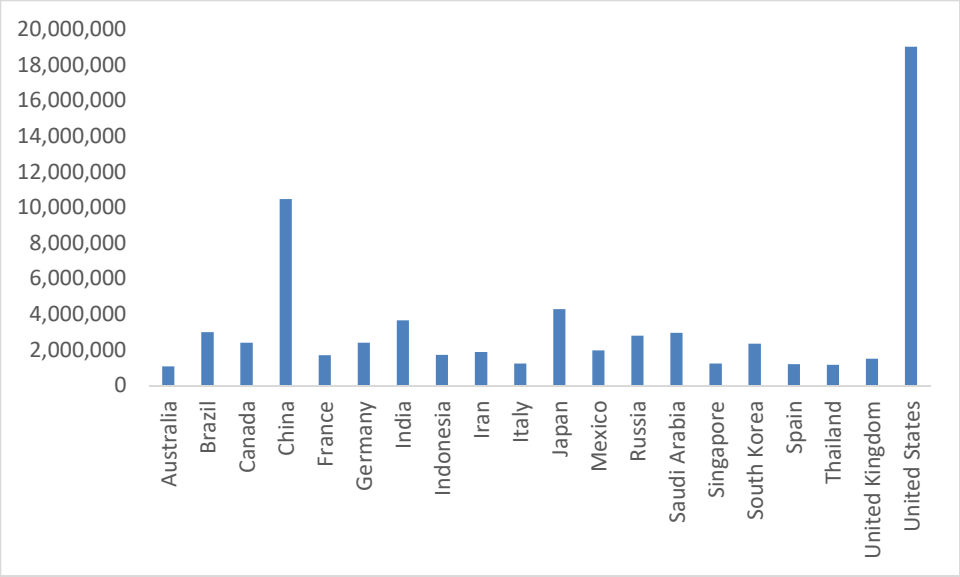
2.13


350,000,000,000
300,000,000,000
250,000,000,000
200,000,000,000
150,000,000,000
100,000,000,000
50,000,000,000
0
Brazil
Canada
China
Iran
Iraq
Kazakhstan
Kuwait
Libya
Nigeria
Qatar
Russia
Saudi Arabia
United Arab Emirates
United States
Venezuela

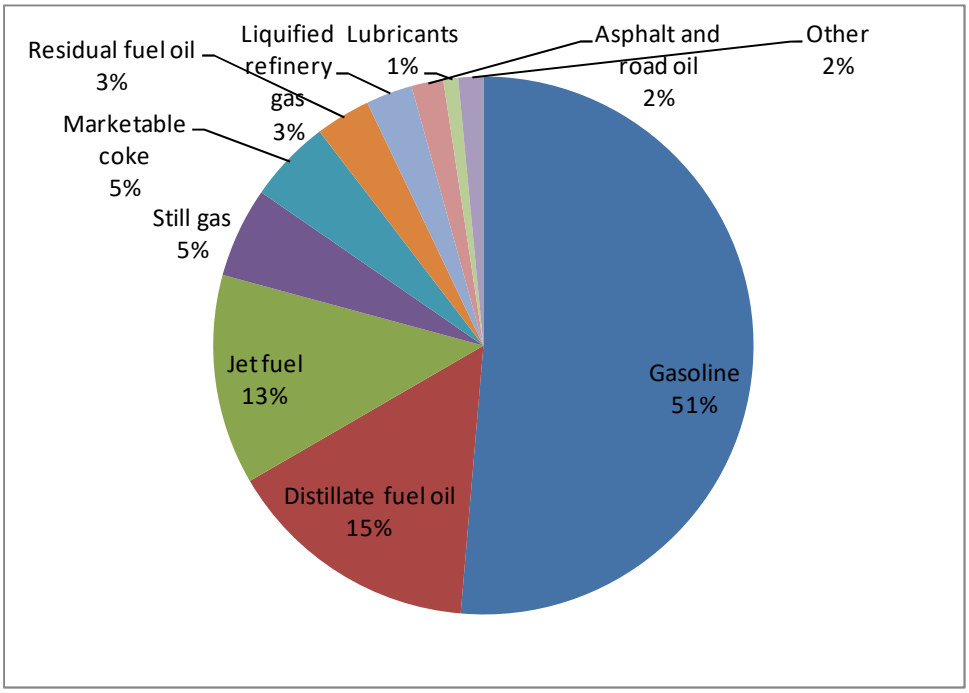
2.14


Percentage
Brazil, 1.0%
Canada, 11.0%
China, 1.6%
Iran, 10.1%
Iraq, 9.2%
Kazakhstan, 1.9%
Kuwait, 6.7%
Libya, 3.1%
Nigeria, 2.4%
Qatar, 1.6%
Russia, 6.6%
Saudi Arabia, 17.2%
United Arab Emirates, 6.3%
United States, 2.3%
Venezuela, 19.1%

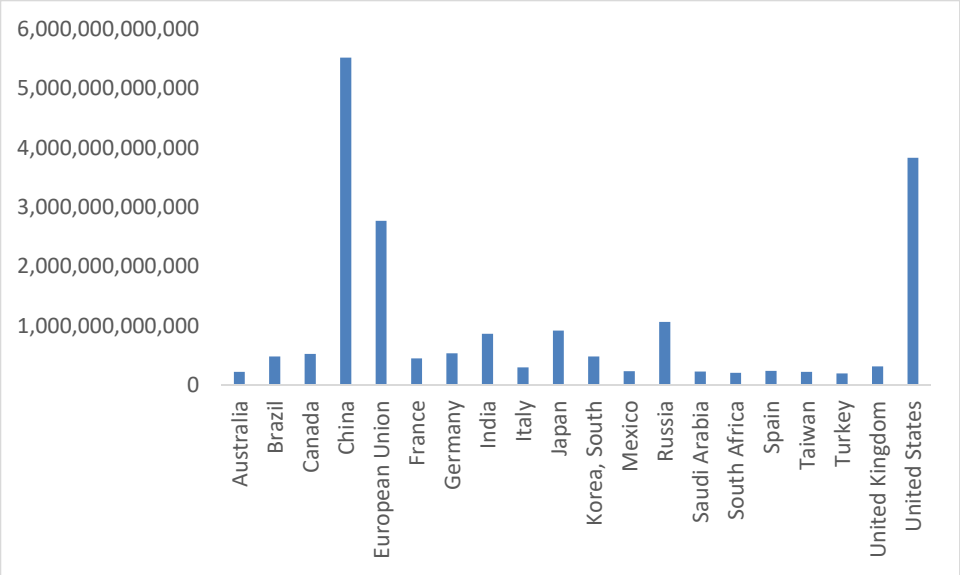
2.15


20,000,000
18,000,000
16,000,000
14,000,000
12,000,000
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
0
Australia
Brazil
Canada
China
France
Germany
India
Indonesia
Iran
Italy
Japan
Mexico
Russia
Saudi Arabia
Singapore
South Korea
Spain
Thailand
United Kingdom
United States

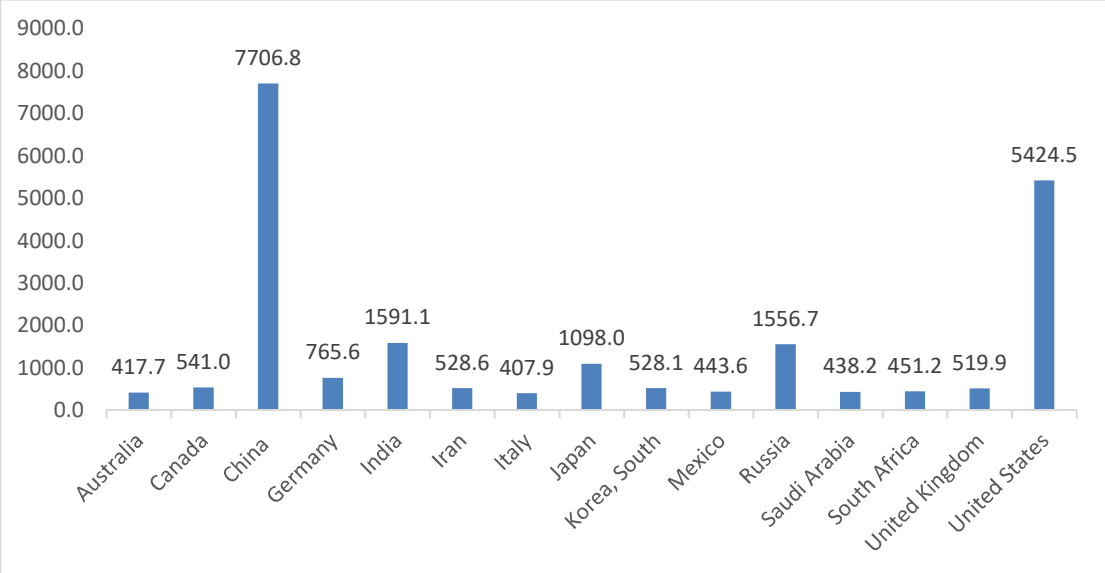
2.16


Residual fuel oil 3%
Liquified refinery gas 3%
Lubricants 1%
Asphalt and road oil 2%
Other 2%
Marketable coke 5%
Still gas 5%
Jet fuel 13%
Gasoline 51%
Distillate fuel oil 15%

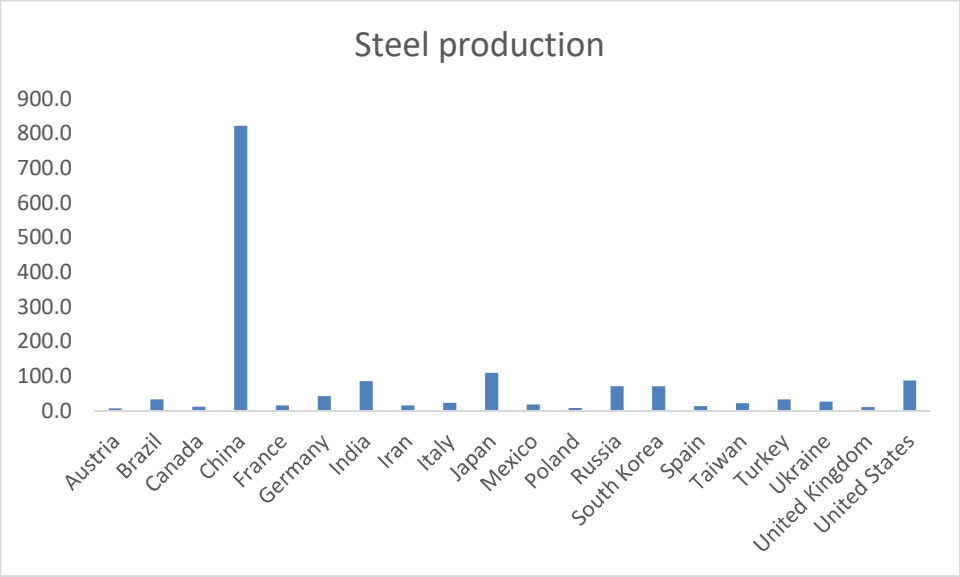
2.17


6,000,000,000,000
5,000,000,000,000
4,000,000,000,000
3,000,000,000,000
2,000,000,000,000
1,000,000,000,000
0
Australia
Brazil
Canada
China
European Union
France
Germany
India
Italy
Japan
Korea, South
Mexico
Russia
Saudi Arabia
South Africa
Spain
Taiwan
Turkey
United Kingdom
United States

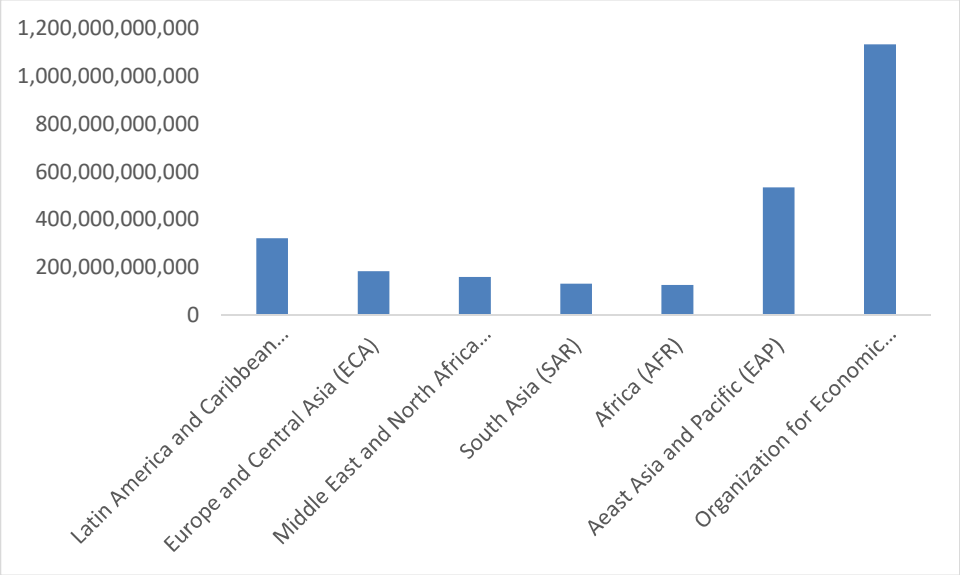
2.18


9000.0
8000.0
7000.0
6000.0
5000.0
4000.0
3000.0
2000.0
1000.0
0.0
417.7
541.0
7706.8
765.6
1591.1
528.6
407.9
1098.0
528.1
443.6
1556.7
438.2
451.2
519.9
5424.5
Australia
Canada
China
Germany
India
Iran
Italy
Japan
Korea, South
Mexico
Russia
Saudi Arabia
South Africa
United Kingdom
United States

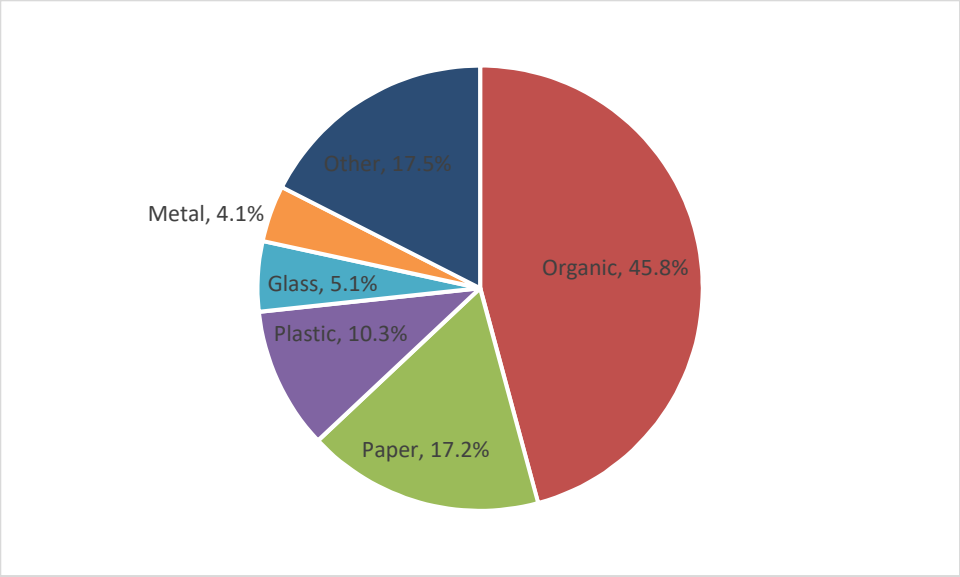
2.19


Steel production
900.0
800.0
700.0
600.0
500.0
400.0
300.0
200.0
100.0
0.0
Austria
Brazil
Canada
China
France
Germany
India
Iran
Italy
Japan
Mexico
Poland
Russia
South Korea
Spain
Taiwan
Turkey
Ukraine
United Kingdom
United States

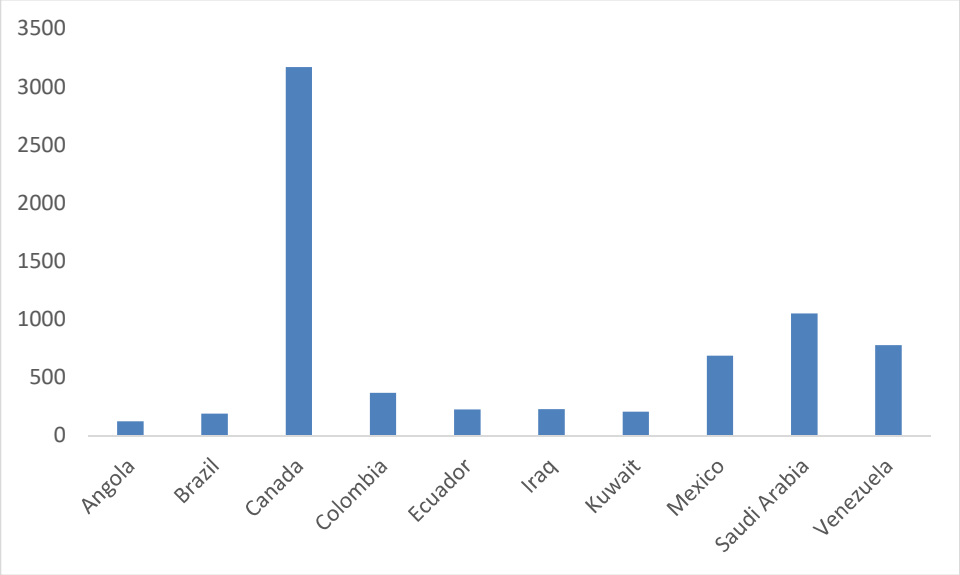
2.20


1,200,000,000,000
1,000,000,000,000
800,000,000,000
600,000,000,000
400,000,000,000
200,000,000,000
0
Latin America and Caribbean...
Europe and Central Asia (ECA)
Middle East and North Africa...
South Asia (SAR)
Africa (AFR)
Aeast Asia and Pacific (EAP)
Organization for Economic...

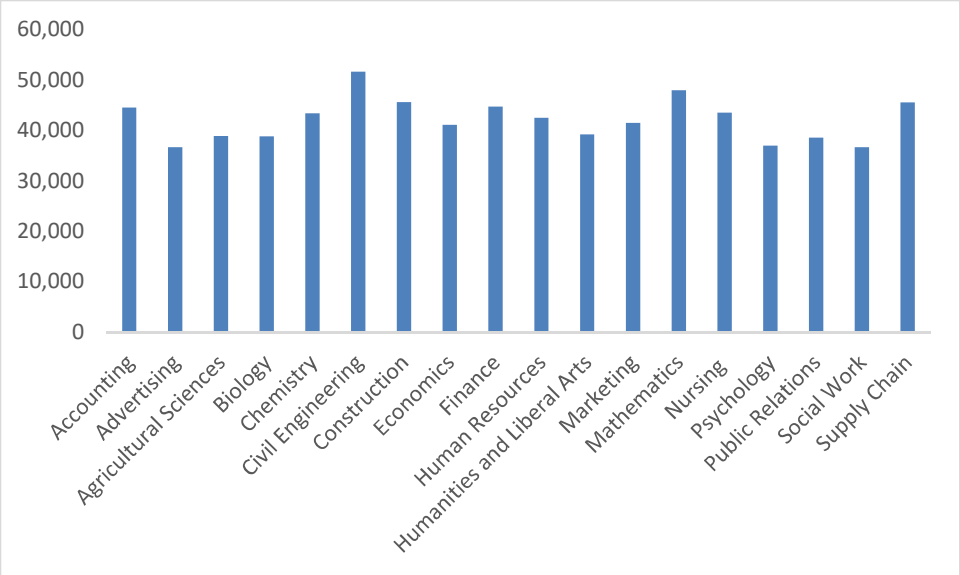
2.21


Other, 17.5%
Metal, 4.1%
Glass, 5.1%
Plastic, 10.3%
Organic, 45.8%
Paper, 17.2%

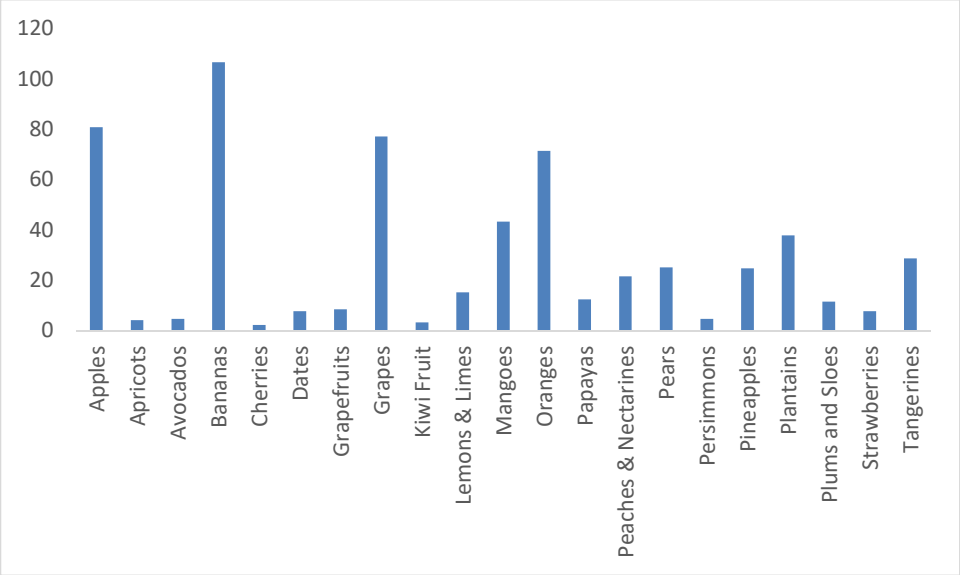
2.22


3500
3000
2500
2000
1500
1000
500
0
Angola
Brazil
Canada
Colombia
Ecuador
Iraq
Kuwait
Mexico
Saudi Arabia
Venezuela

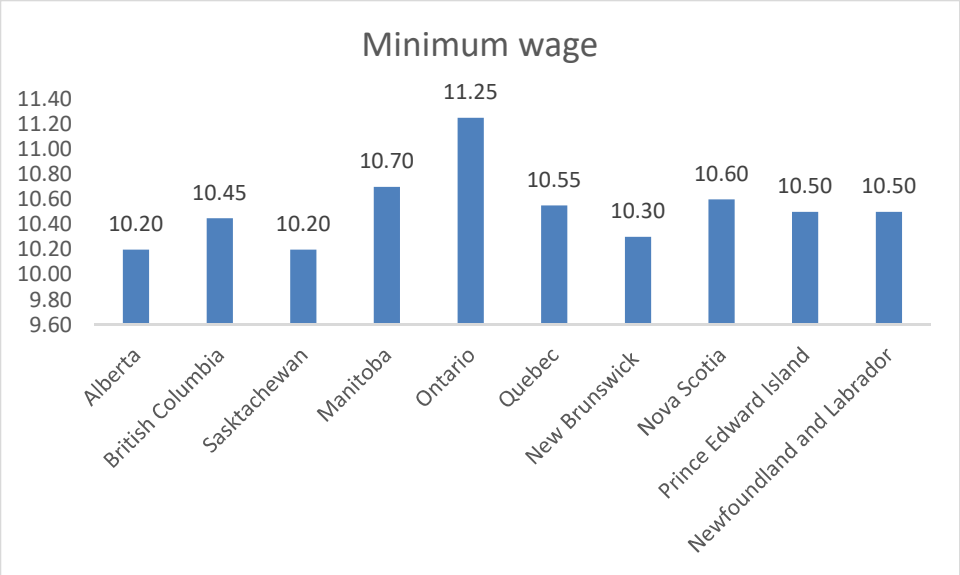
2.23

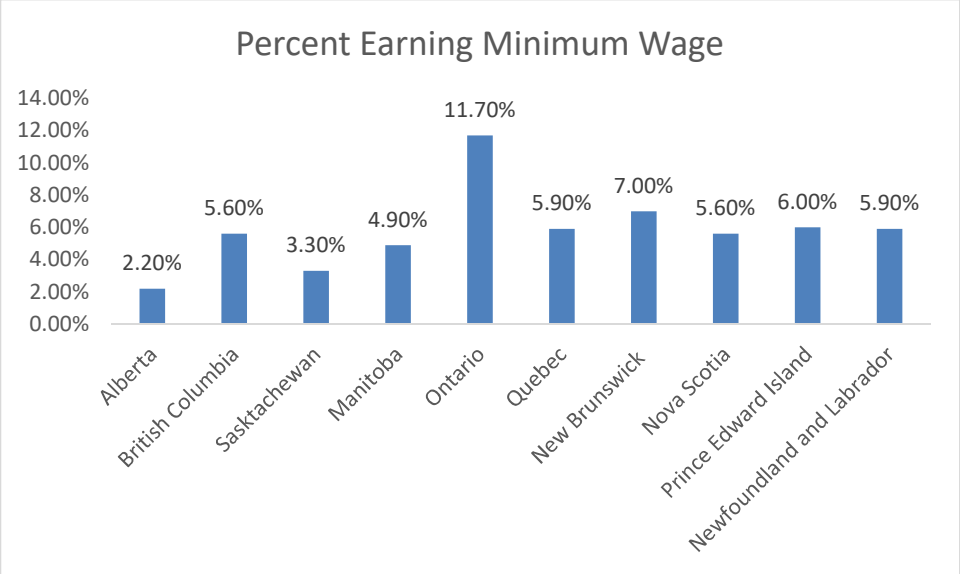

60,000
50,000
40,000
30,000
20,000
10,000
0
Accounting
Advertising
Agricultural Sciences
Biology
Chemistry
Civil Engineering
Construction
Economics
Finance
Human Resources
Humanities and Liberal Arts
Marketing
Mathematics
Nursing
Psychology
Public Relations
Social Work
Supply Chain

2.24

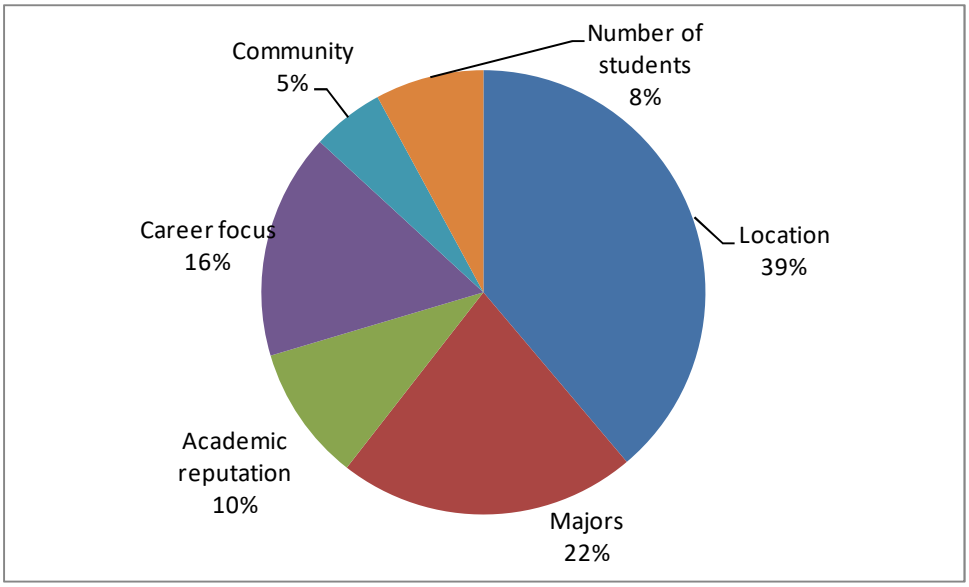

120
100
80
60
40
20
0
Apples
Apricots
Avocados
Bananas
Cherries
Dates
Grapefruits
Grapes
Kiwi Fruit
Lemons & Limes
Mangoes
Oranges
Papayas
Peaches & Nectarines
Pears
Persimmons
Pineapples
Plantains
Plums and Sloes
Strawberries
Tangerines

2.25

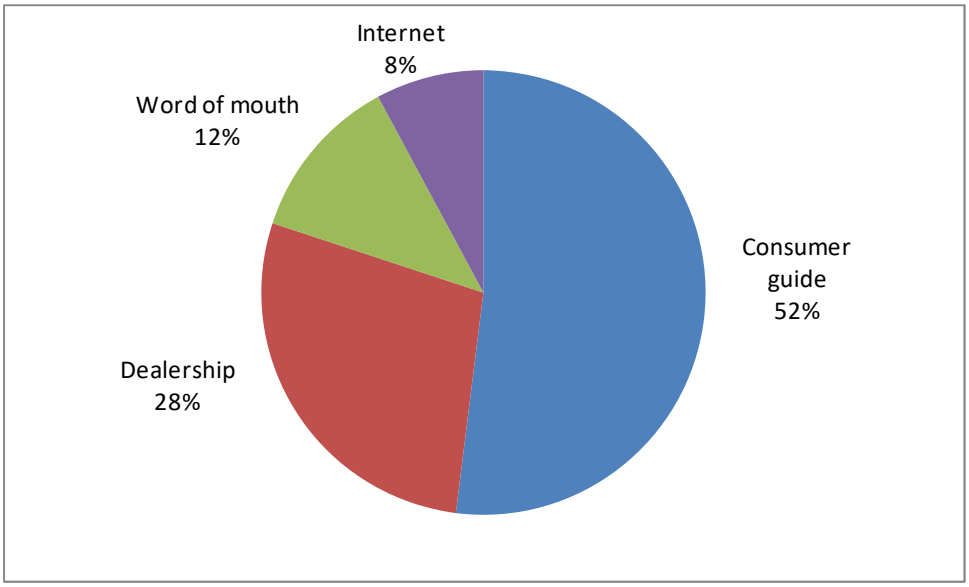

Minimum wage
11.40
11.20
11.00
10.80
10.60
10.40
10.20
10.00
9.80
9.60
10.20
10.45
10.20
10.70
11.25
10.55
10.30
10.60
10.50
10.50
Alberta
British Columbia
Sasktachewan
Manitoba
Ontario
Quebec
New Brunswick
Nova Scotia
Prince Edward Island
Newfoundland and Labrador


Percent Earning Minimum Wage
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
2.20%
5.60%
3.30%
4.90%
11.70%
5.90%
7.00%
5.60%
6.00%
5.90%
Alberta
British Columbia
Sasktachewan
Manitoba
Ontario
Quebec
New Brunswick
Nova Scotia
Prince Edward Island
Newfoundland and Labrador

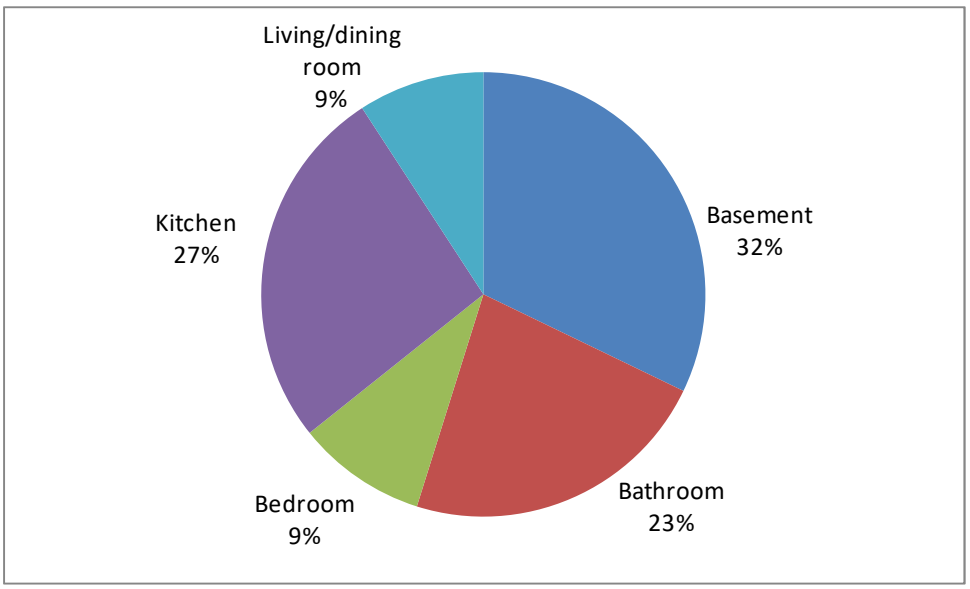
2.26


Community
5%
Number of
students
8%
Career focus
16%
Location
39%
Academic
reputation
10%
Majors
22%

2.27


Internet
8%
Word of mouth
12%
Consumer
guide
52%
Dealership
28%

2.28

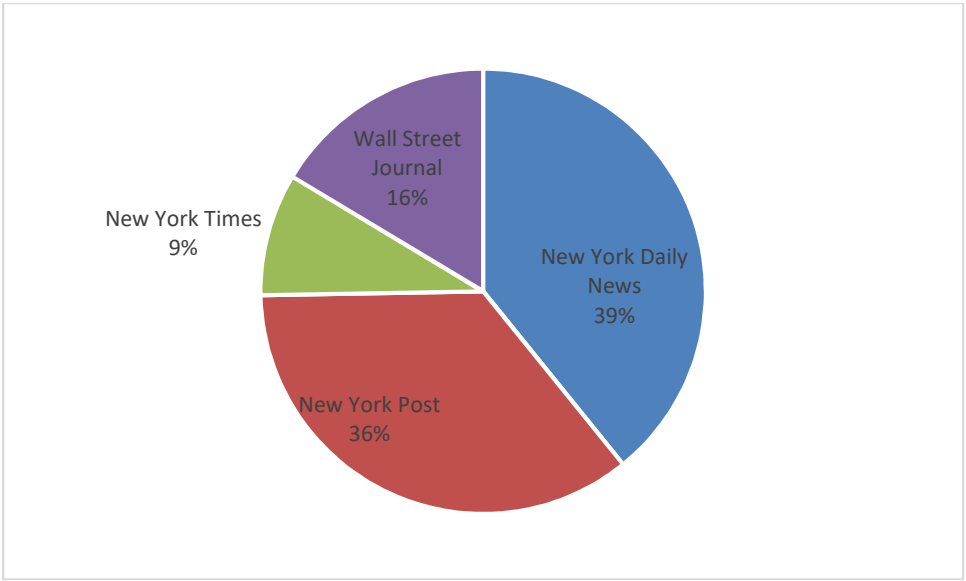


The basement is the top choice followed by kitchen, bathroom, bedroom, and living/dining room.

2.29 a

Newspaper	Frequency	Relative Frequency
Daily News	141	39.2%
Post	128	35.6%
Times	32	8.9%
WSJ	59	16.4%

b

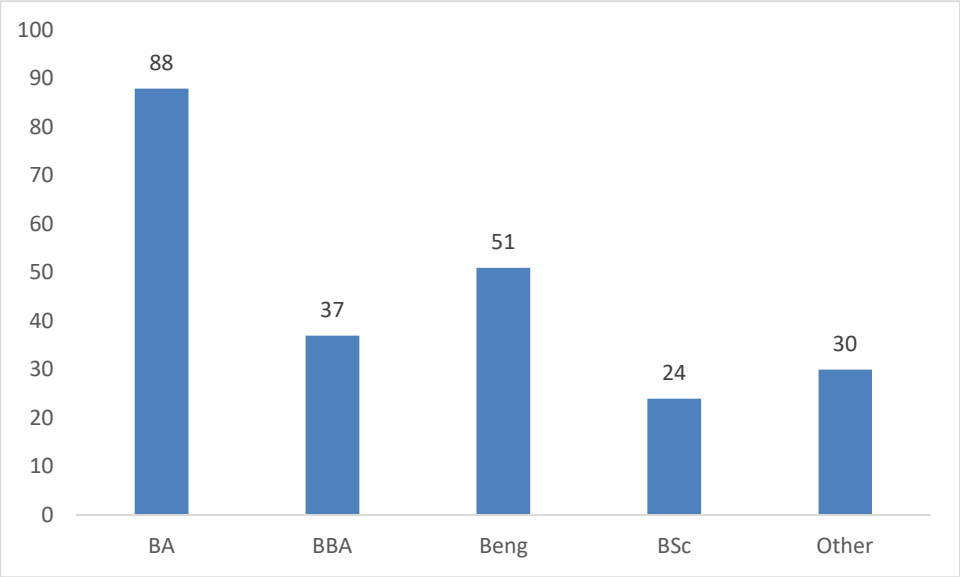


The Daily News and the Post dominate the market

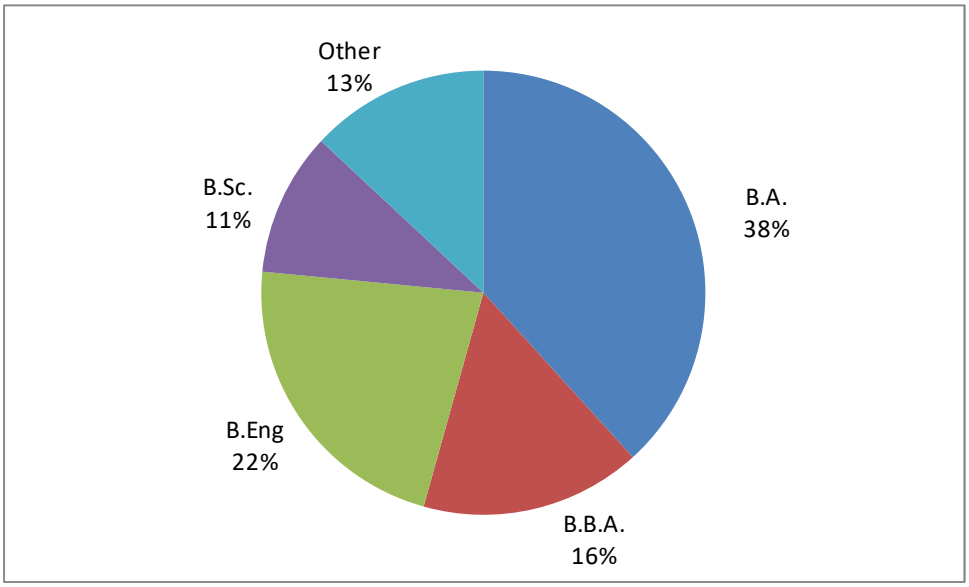
2.30a

Degree	Frequency
BA	88
BBA	37
B Eng	51
B Sc	24
Other	30

b.

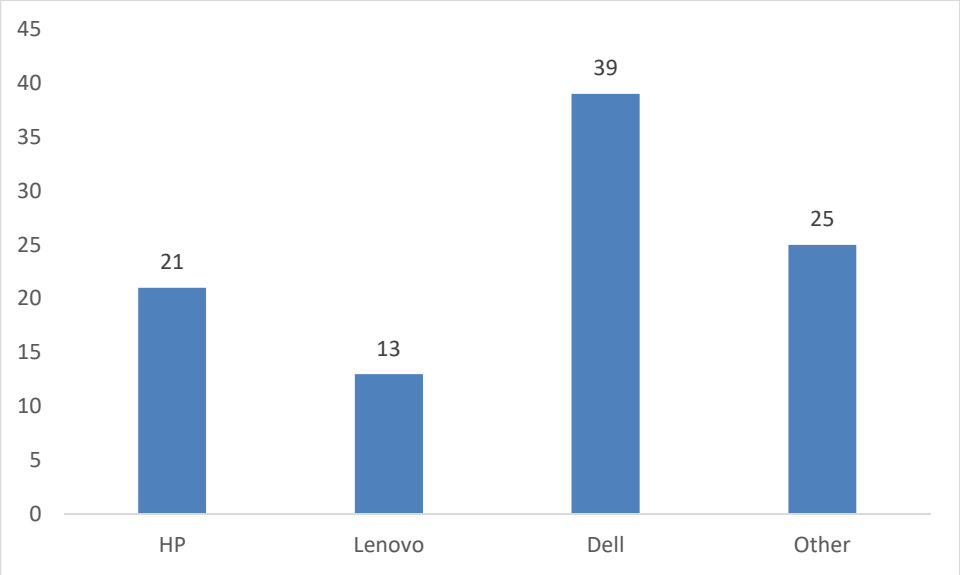


c

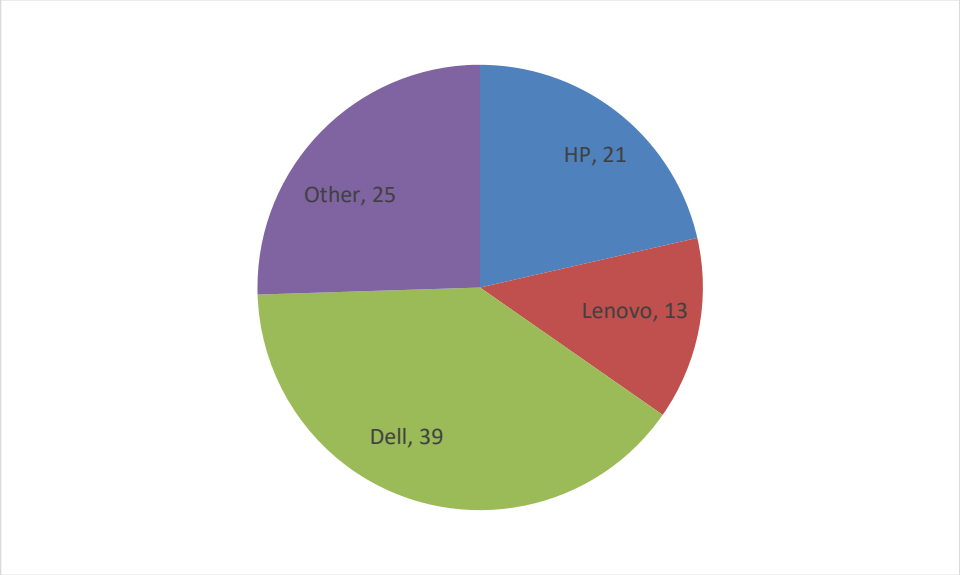


d. About 4 applicants in 10 have the BA degree, about one-fifth have a BEng. and one-sixth have a BBA.

2.31a



b

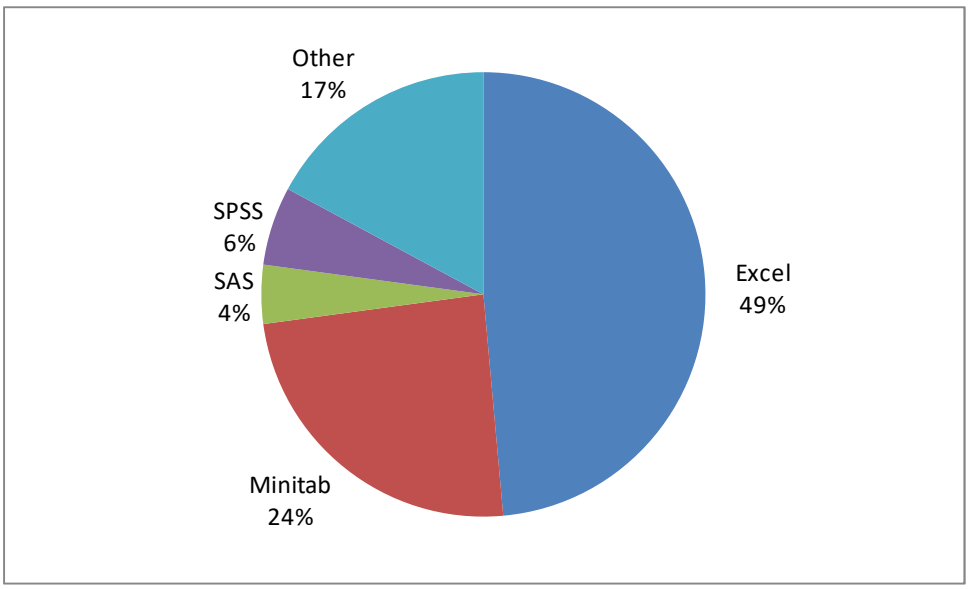


c Dell is most popular with 40% proportion, followed by other, 26%, HP, 21% and Lenovo, 13%.

2.32 a

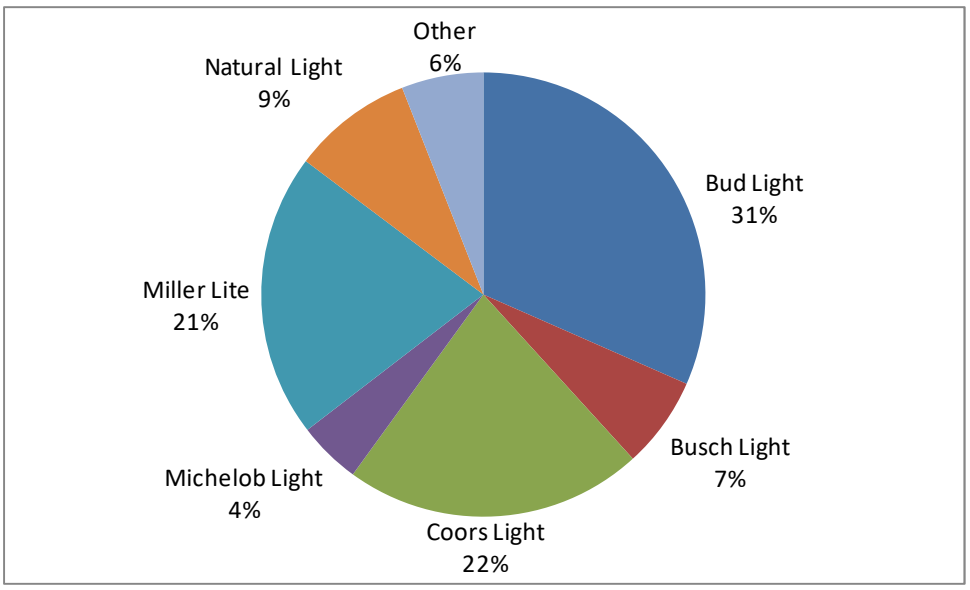
Software	Frequency
Excel	34
Minitab	17
SAS	3
SPSS	4
Other	12

b

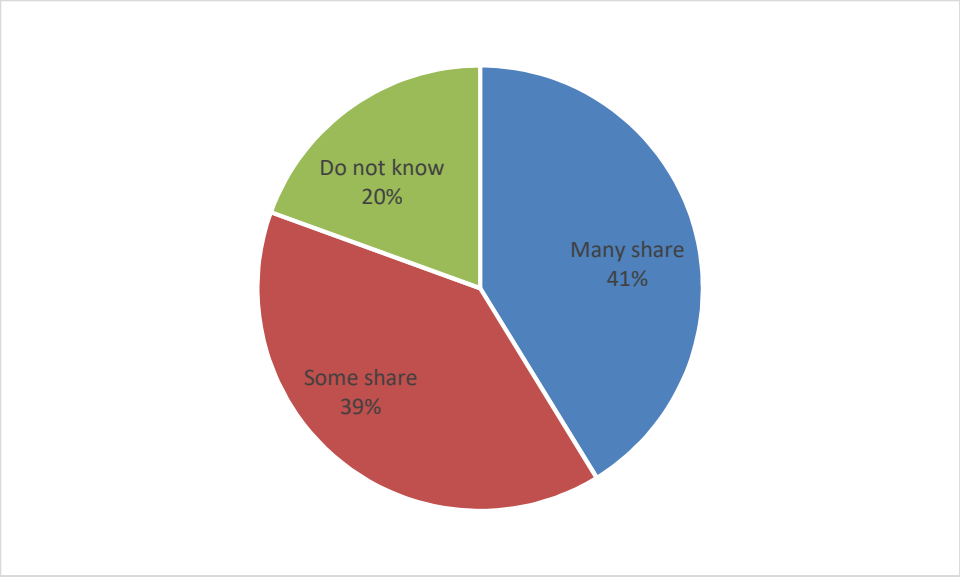


c Excel is the choice of about half the sample, one-quarter have opted for Minitab, and a small fraction chose SAS and SPSS.

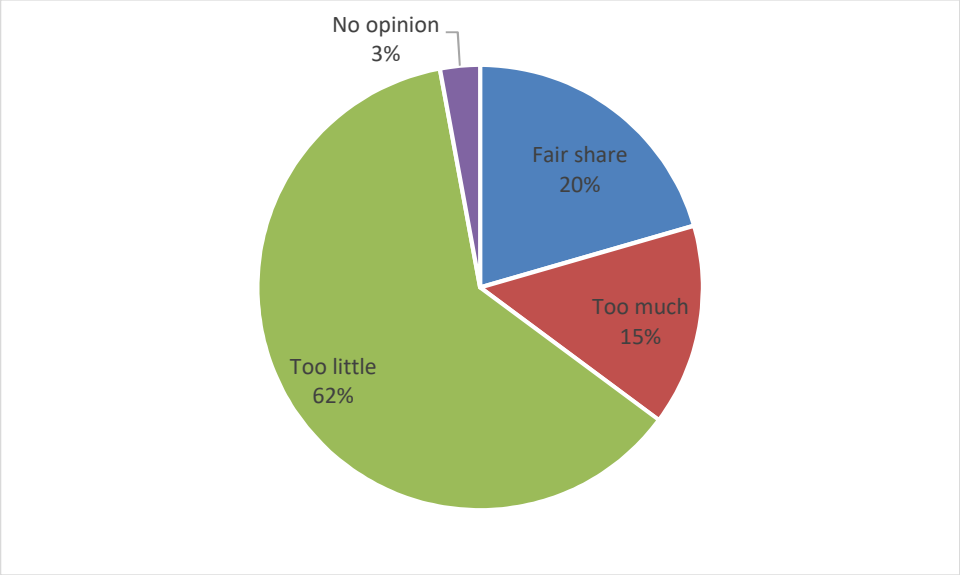
2.33



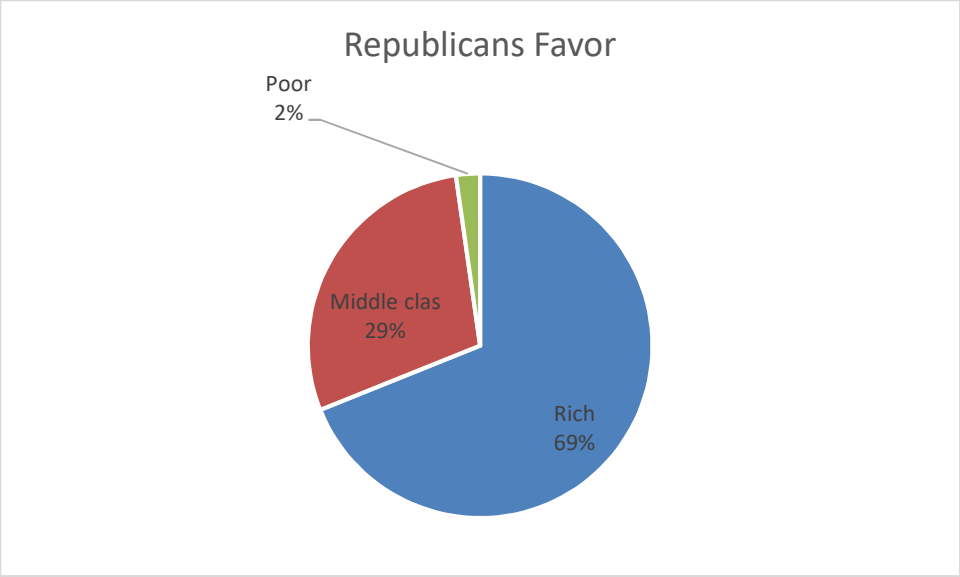
2.34


Do not know
20%
Many share
41%
Some share
39%

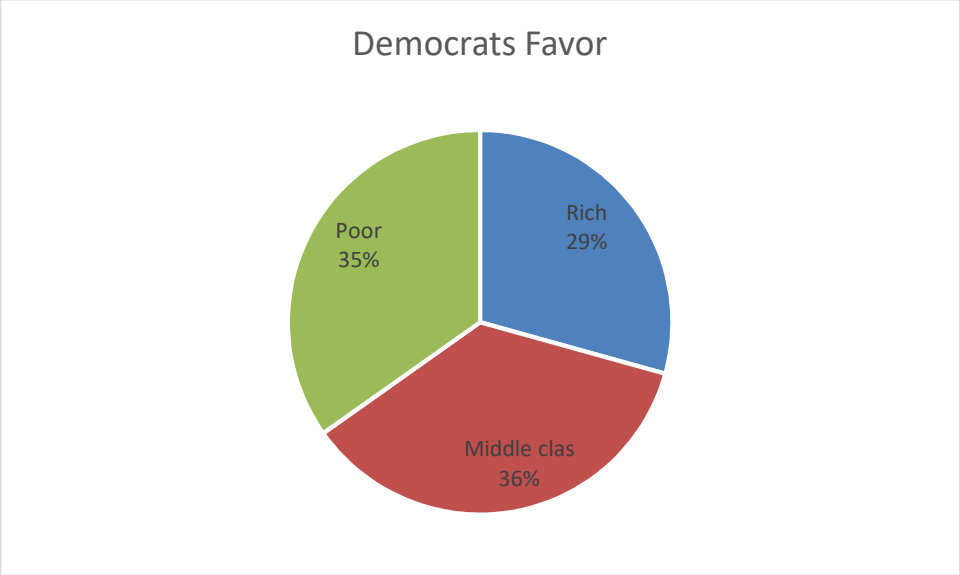
2.35


No opinion
3%
Fair share
20%
Too much
15%
Too little
62%

2.36 a



b

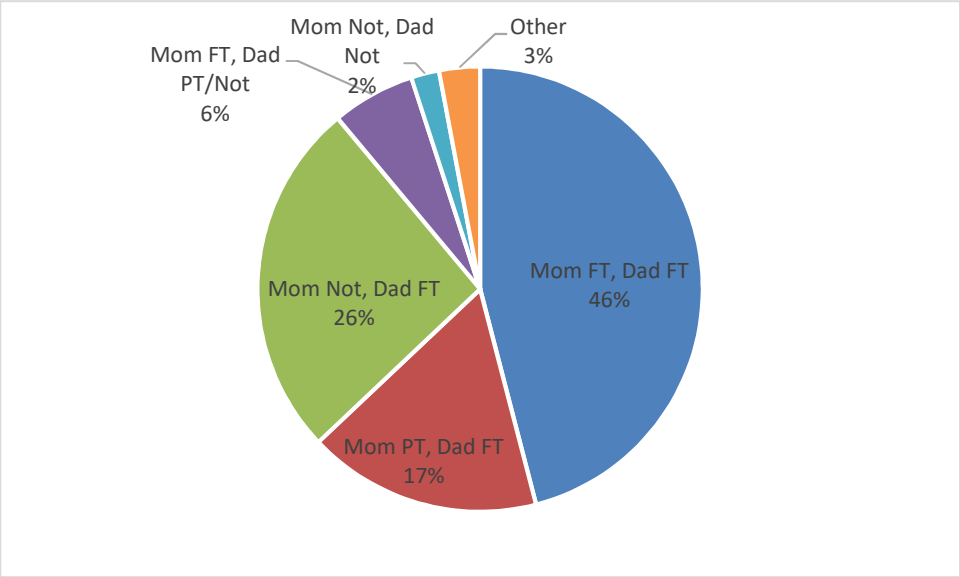


According to the survey Republicans favor the rich and Democrats are split among the middle class, poor, and rich.

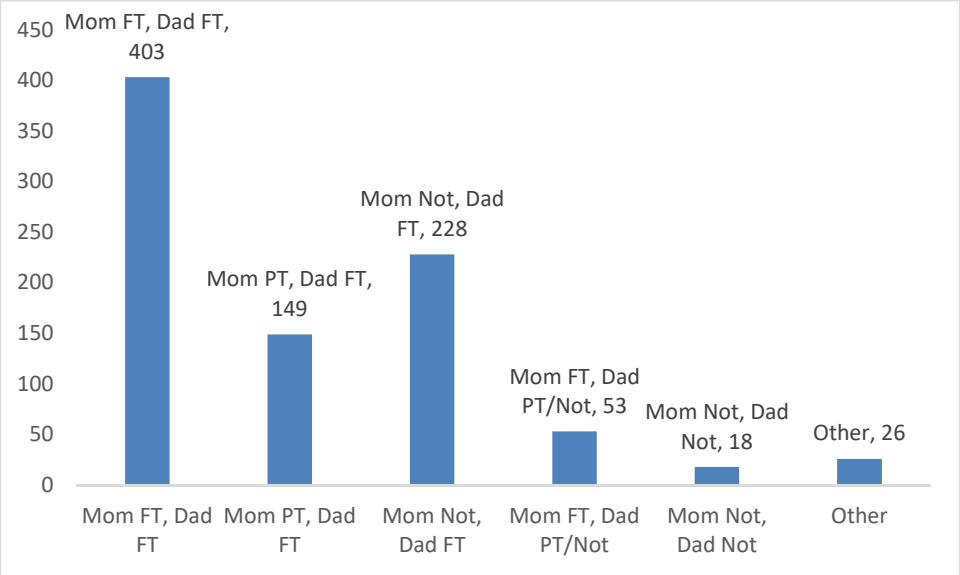
2.37 a

Category	Frequency	Relative Frequency
Mom: Full time, Dad: Full time	403	46.0%
Mom: Part time, Dad: Full time	149	17.0%
Mom: Not employed, Dad: Full time	228	26.0%
Mom: Full time, Dad: Part time or not employed	53	6.0%
Mom: Not employed, Dad: Not employed	18	2.1%
Other	26	3.0%

b

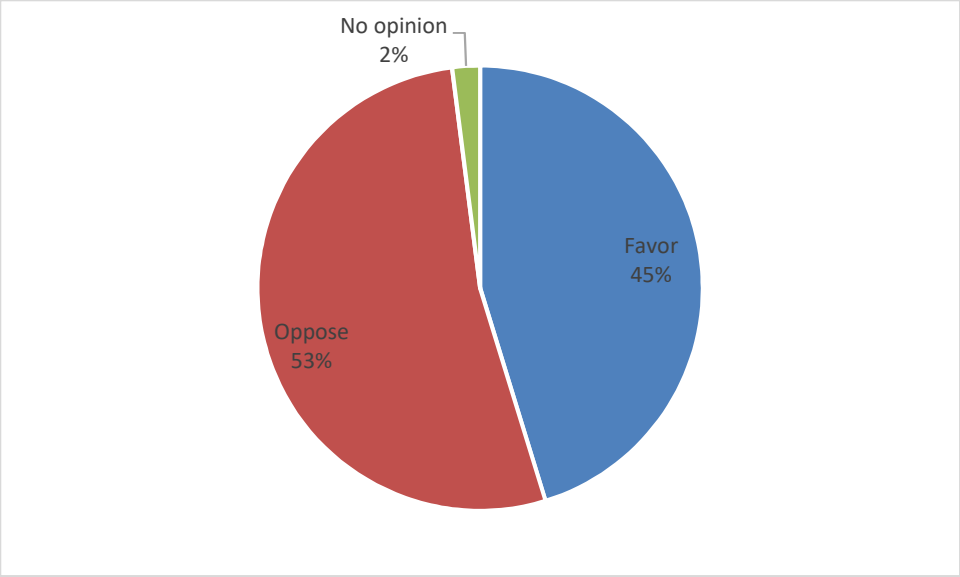


c



d In most households Dad is working full time. There are very few households where neither Mom nor Dad are working.

2.38



A small majority oppose the Affordable Care Act.

2.39a

Views on social issues	Frequency	Relative Frequency
Liberal	322	31.4%
Moderate	328	32.0%
Conservative	375	36.6%

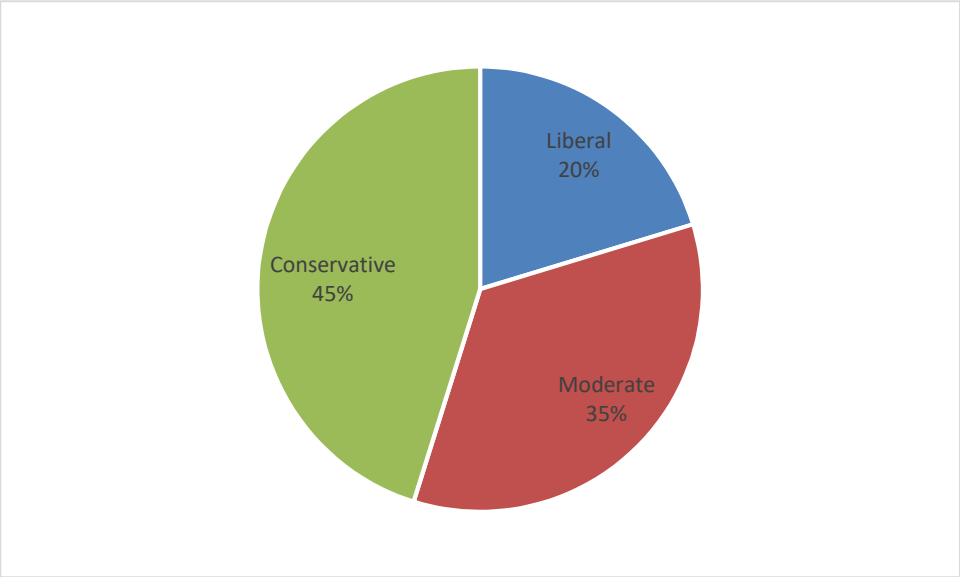
b

The country is split among the three views on social issues with a small plurality of conservatives.

2.40 a

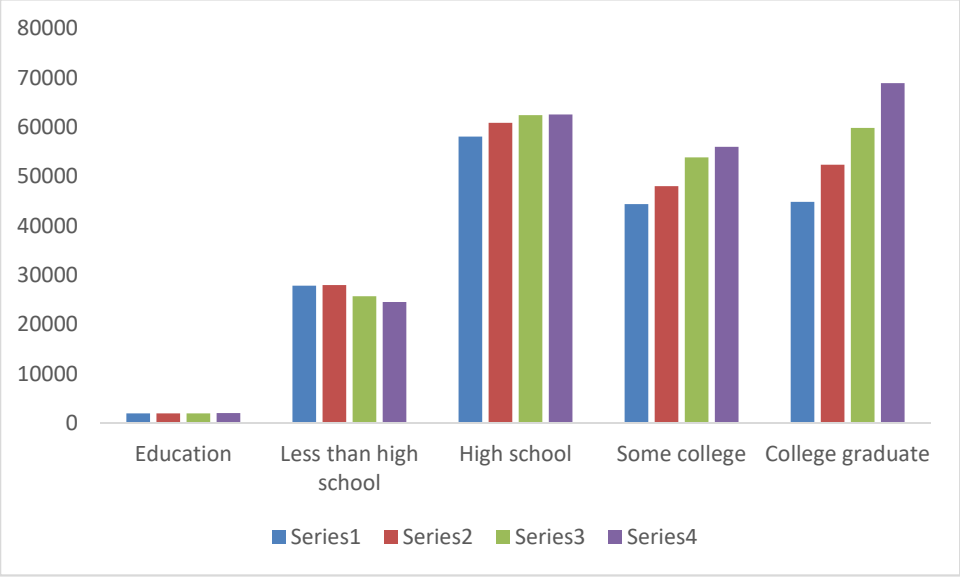
Views on economic issues	Frequency	Relative Frequency
Liberal	208	20.3%
Moderate	354	34.5%
Conservative	463	45.2%

b



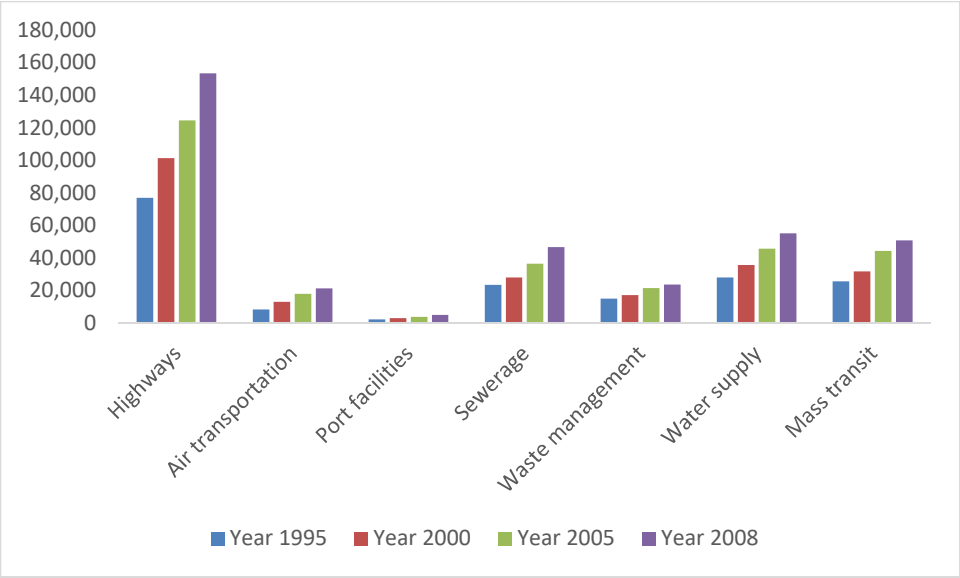
Economically the country is conservative.

2.41



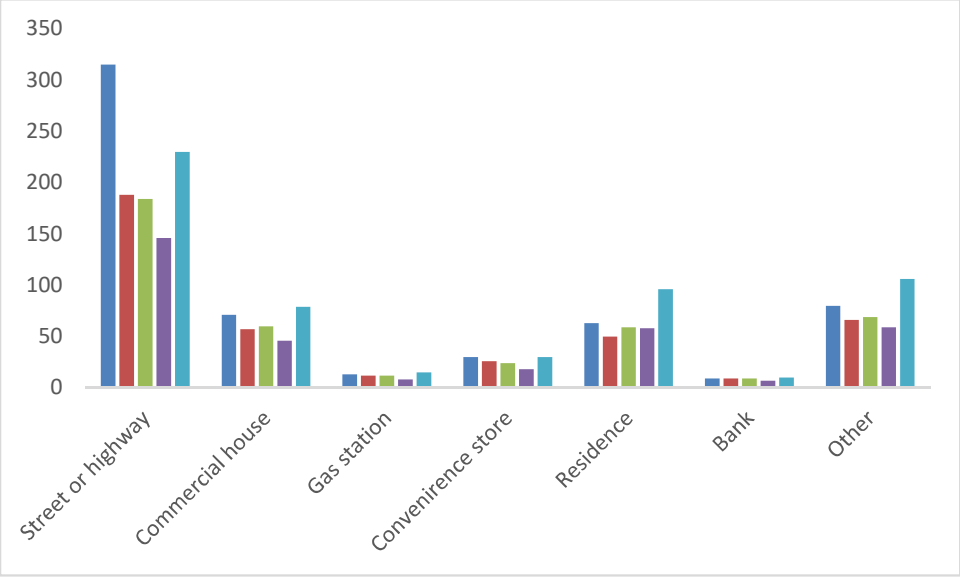
There is decreasing numbers of Americans who did not finish high school and increasing numbers of those that go to college.

2.42



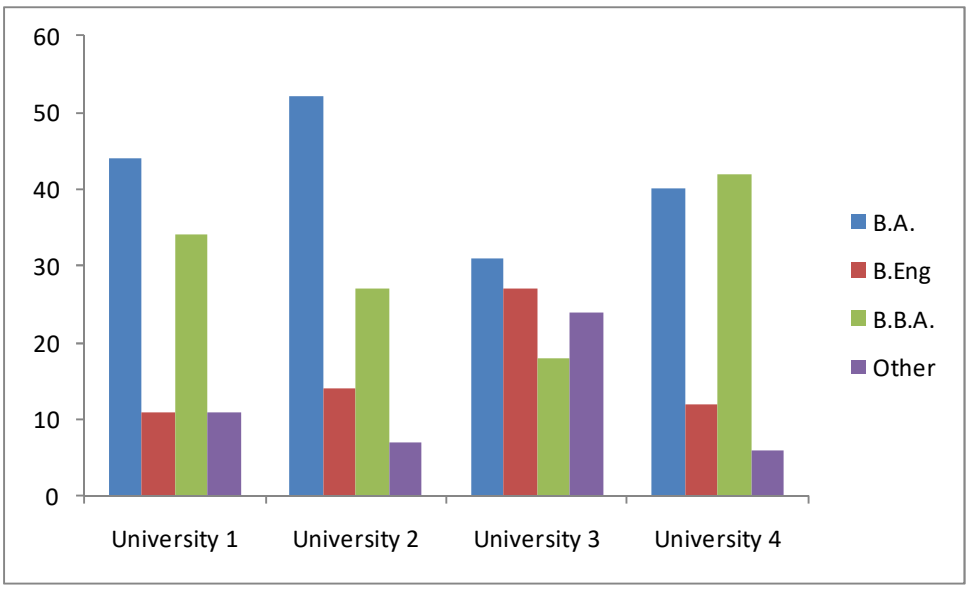
Spending is increasing in all seven areas.

2.43



In general crime was decreasing until 2014 when it started increasing.

2.44



Universities 1 and 2 are similar and quite dissimilar from universities 3 and 4, which also differ. The two nominal variables appear to be related.

2.45

Count of Owner	Last				
Second-last	Exxon	Amoco	Texaco	Other	Grand Total
Exxon	39	36	51	23	149
Amoco	36	32	46	20	134
Texaco	54	46	65	29	194
Other	24	20	28	10	82
Grand Total	153	134	190	82	559

Count of Owner	Last				
Second-last	Exxon	Amoco	Texaco	Other	Grand Total
Exxon	25%	27%	27%	28%	27%
Amoco	24%	24%	24%	24%	24%
Texaco	35%	34%	34%	35%	35%
Other	16%	15%	15%	12%	15%
Grand Total	100%	100%	100%	100%	100%

The column proportions are similar; the two nominal variables appear to be unrelated. There does not appear to be any brand loyalty.

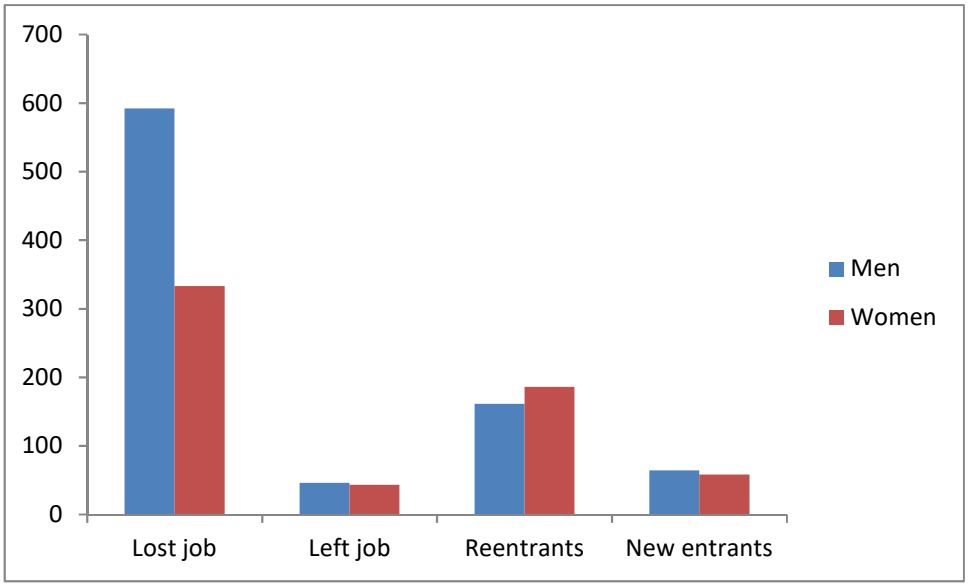
2.46

3	Count of Respondent	Smoke?		
4	Parent	Smoke?	Do not smo	Grand Total
5	Neither	73	14	87
6	Father	26	12	38
7	Mother	31	18	49
8	Both	10	41	51
9	Grand Total	140	85	225

3	Count of Respondent	Smoke?		
4	Parent	Smoke?	Do not smoke	Grand Total
5	Neither	52%	16%	39%
6	Father	19%	14%	17%
7	Mother	22%	21%	22%
8	Both	7%	48%	23%
9	Grand Total	100%	100%	100%

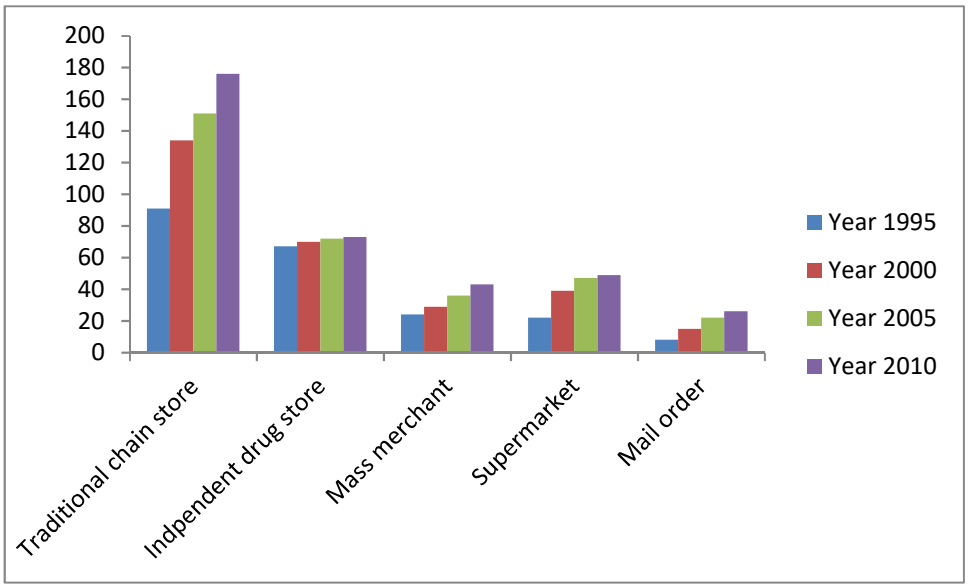
The two variables are related.

2.47



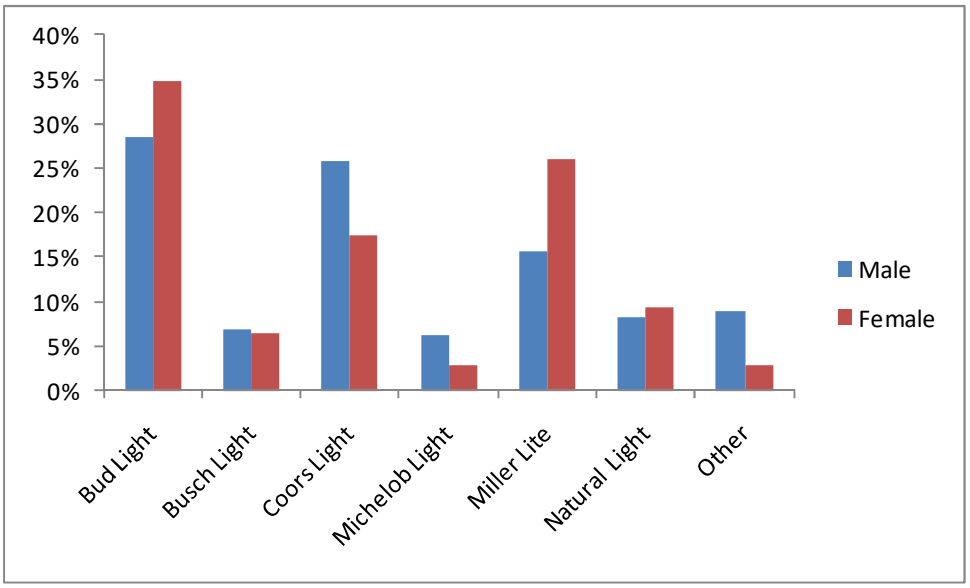
There are large differences between men and women in terms of the reason for unemployment.

2.48



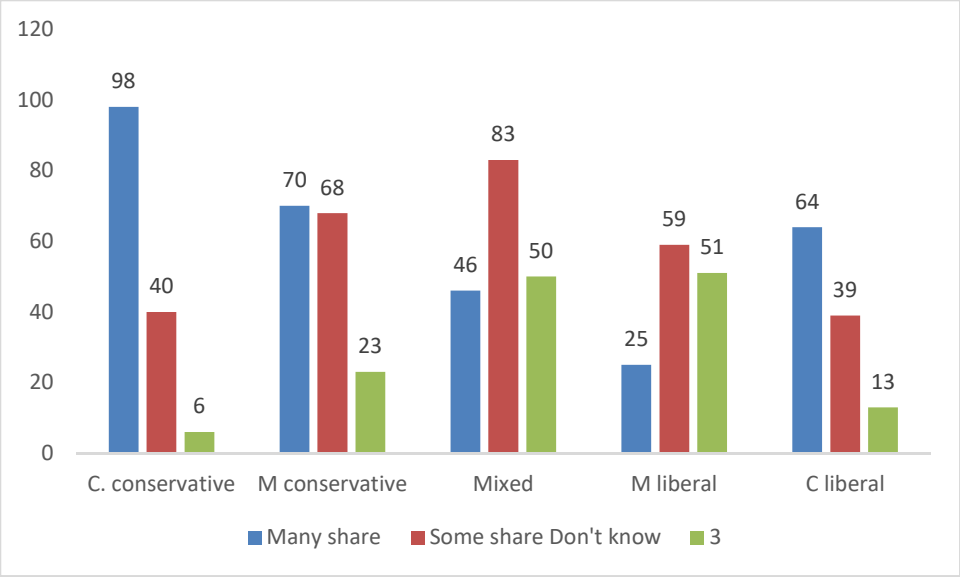
The number of prescriptions filled by all stores except independent drug stores has increased substantially.

2.49



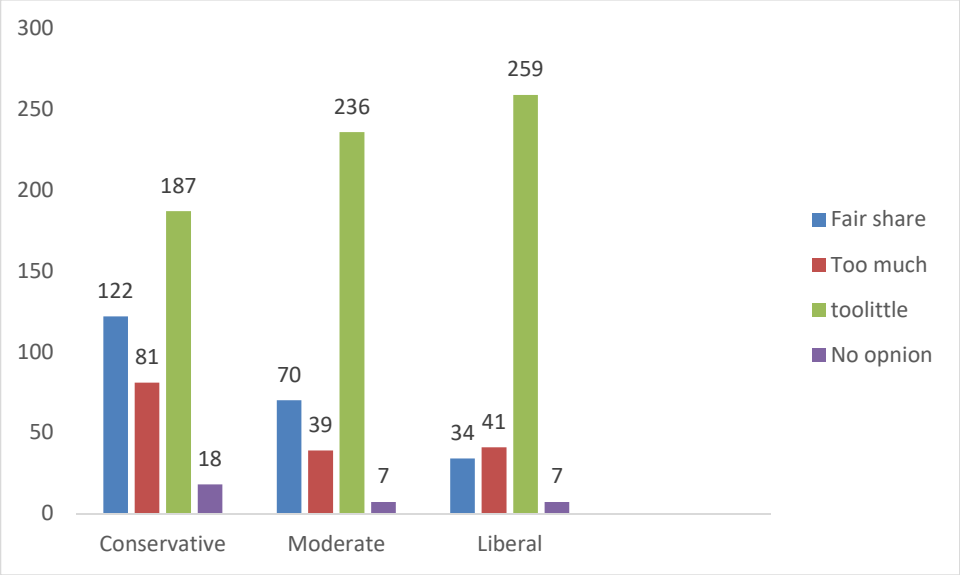
There appears to be differences between female and male students in their choice of light beer.

2.50



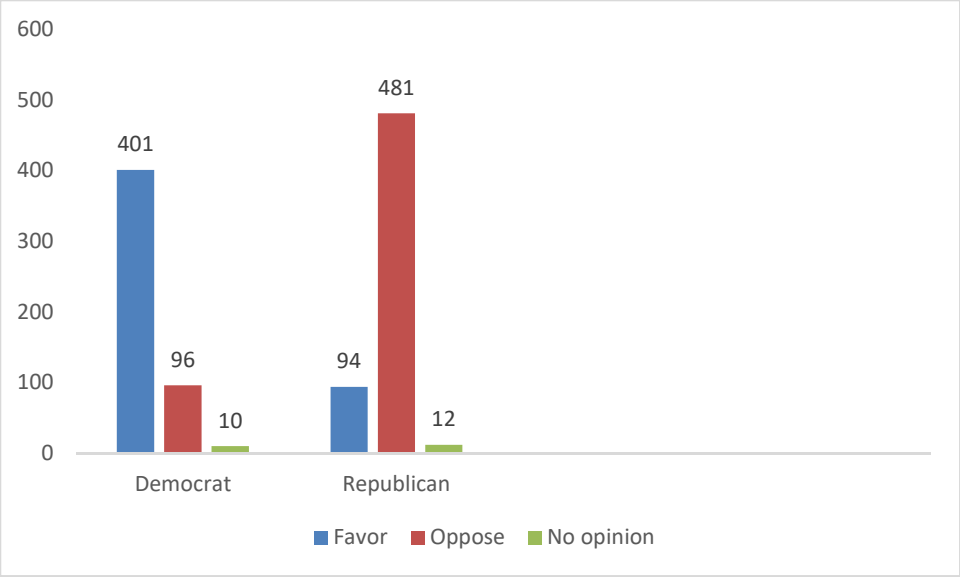
There are differences among the five groups.

2.51



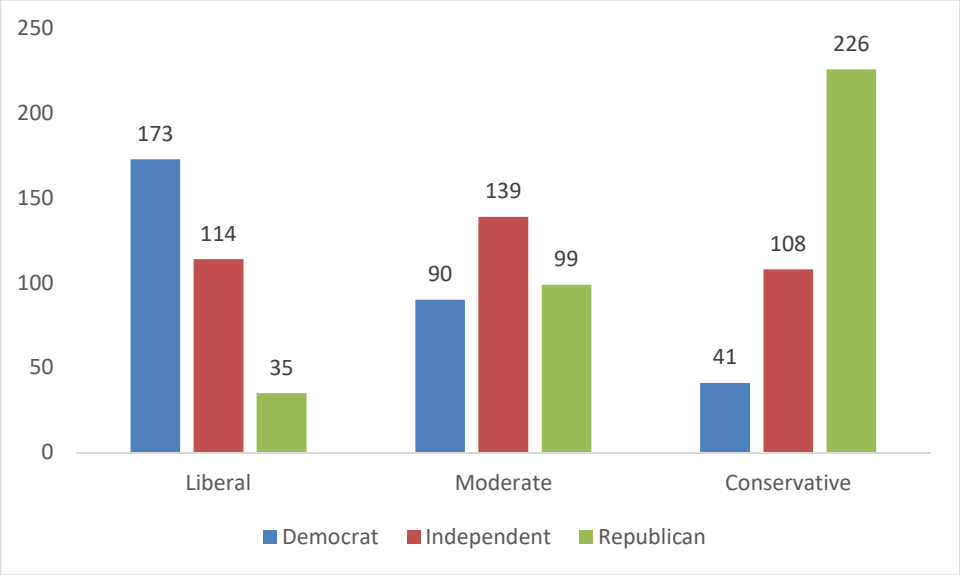
All three groups say that upper-income people pay too little. However Conservatives are more likely to say fair share than Moderates or Liberals

2.52



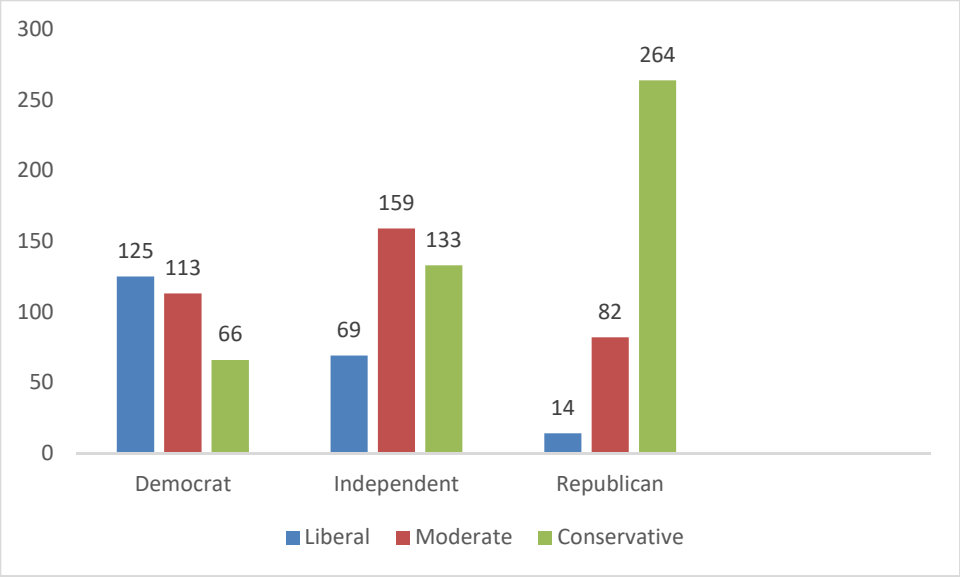
Democrats support and Republicans oppose the Affordable Care Act.

2.53



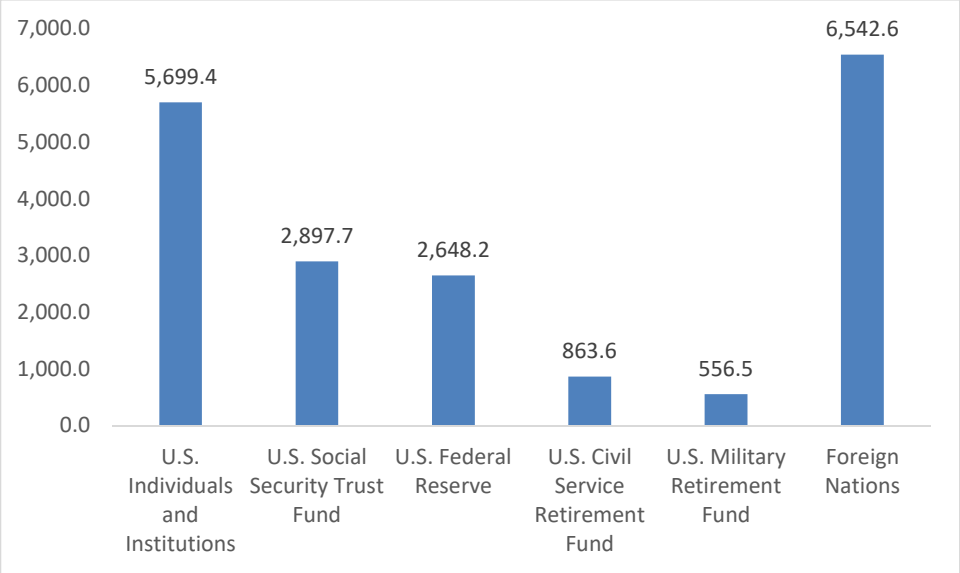
No surprise-on social issues Democrats are liberal and Republicans are conservative.

2.54

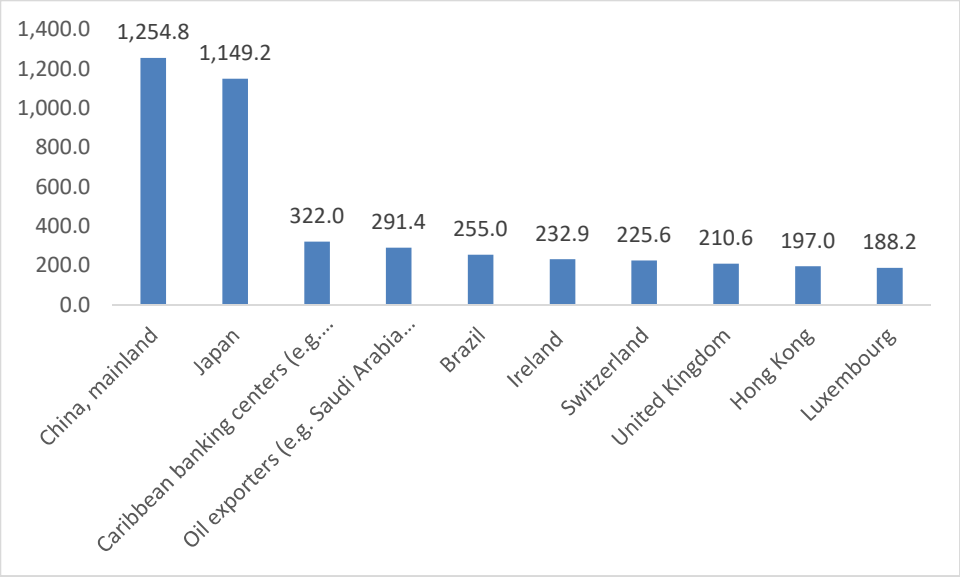


On economic issues Republicans are very conservative whereas Democrats and Moderates are mixed.

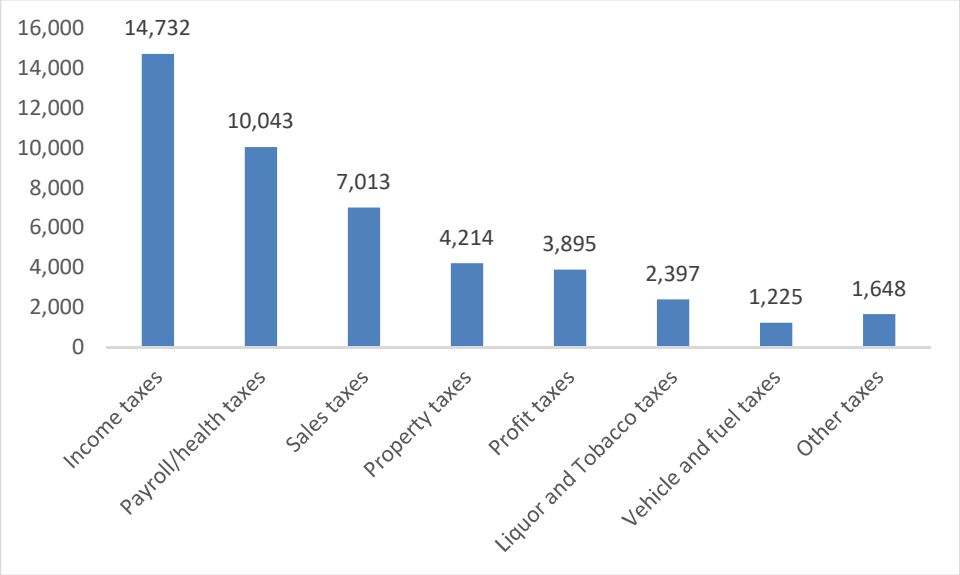
2.55



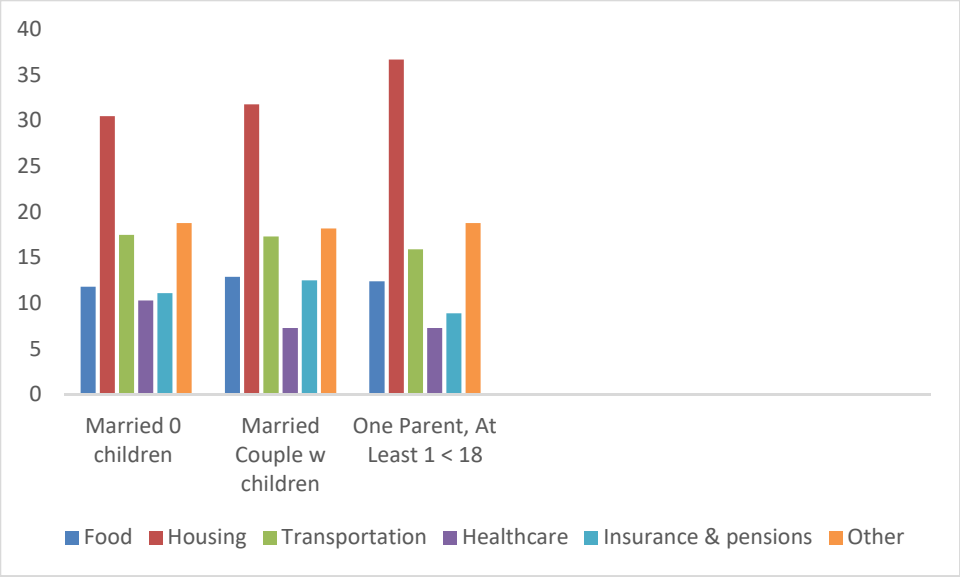
2.56


1,400.0
1,200.0
1,000.0
800.0
600.0
400.0
200.0
0.0
1,254.8
1,149.2
322.0
291.4
255.0
232.9
225.6
210.6
197.0
188.2
China, mainland
Japan
Caribbean banking centers (e.g....
Oil exporters (e.g. Saudi Arabia...
Brazil
Ireland
Switzerland
United Kingdom
Hong Kong
Luxembourg

2.57

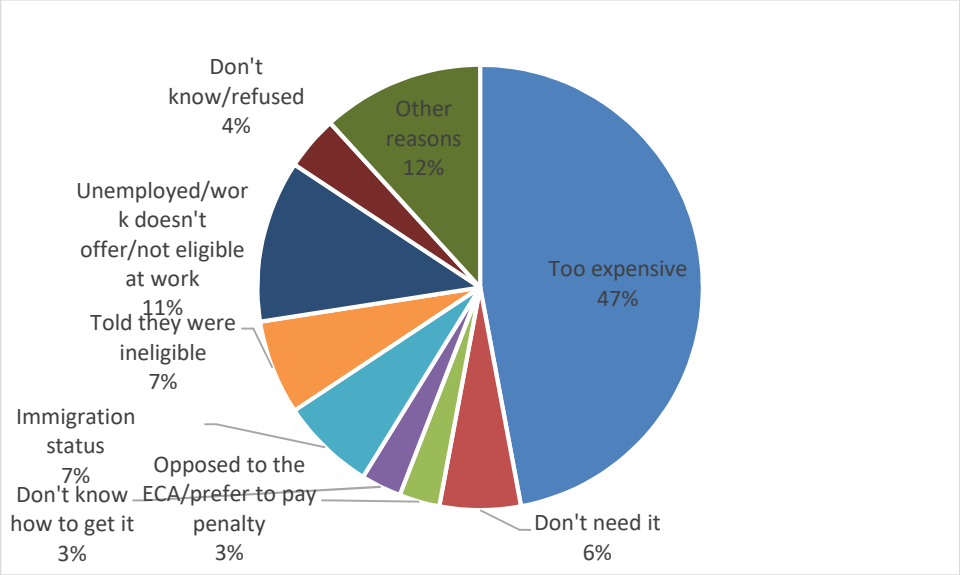

16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
14,732
10,043
7,013
4,214
3,895
2,397
1,225
1,648
Income taxes
Payroll/health taxes
Sales taxes
Property taxes
Profit taxes
Liquor and Tobacco taxes
Vehicle and fuel taxes
Other taxes

2.58

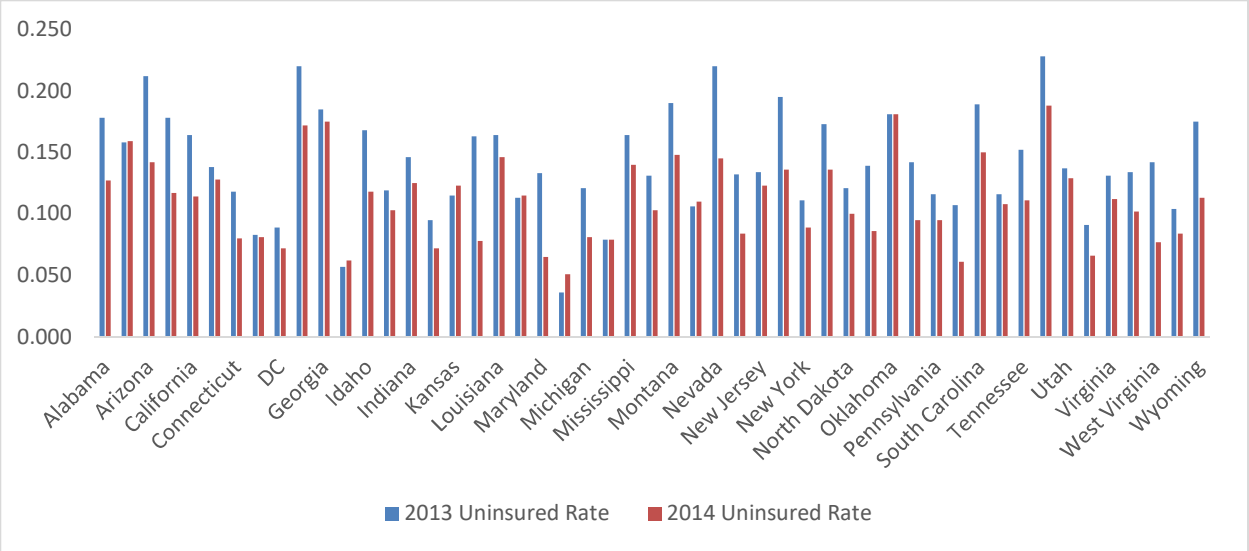


The pattern is about the same for the three households.

2.59

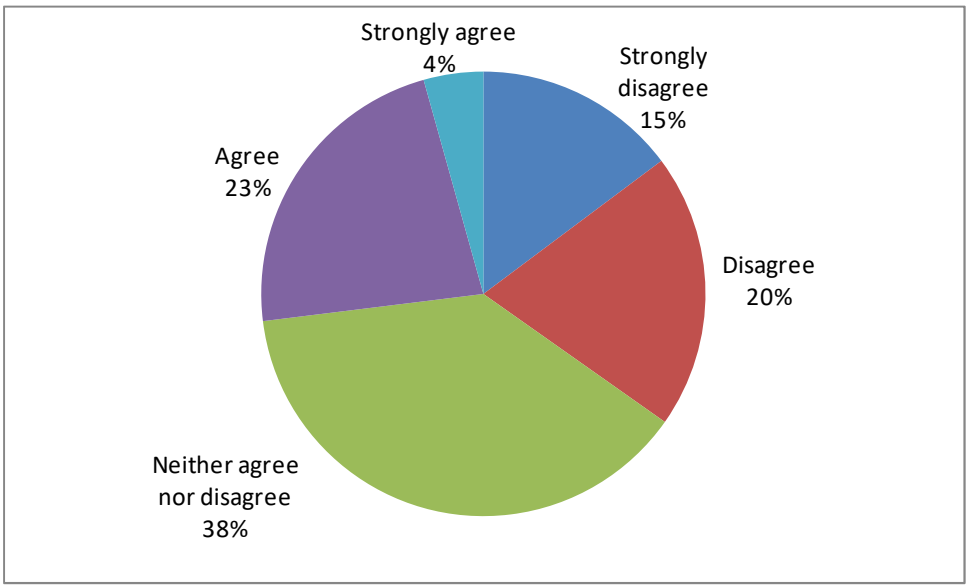


2.60



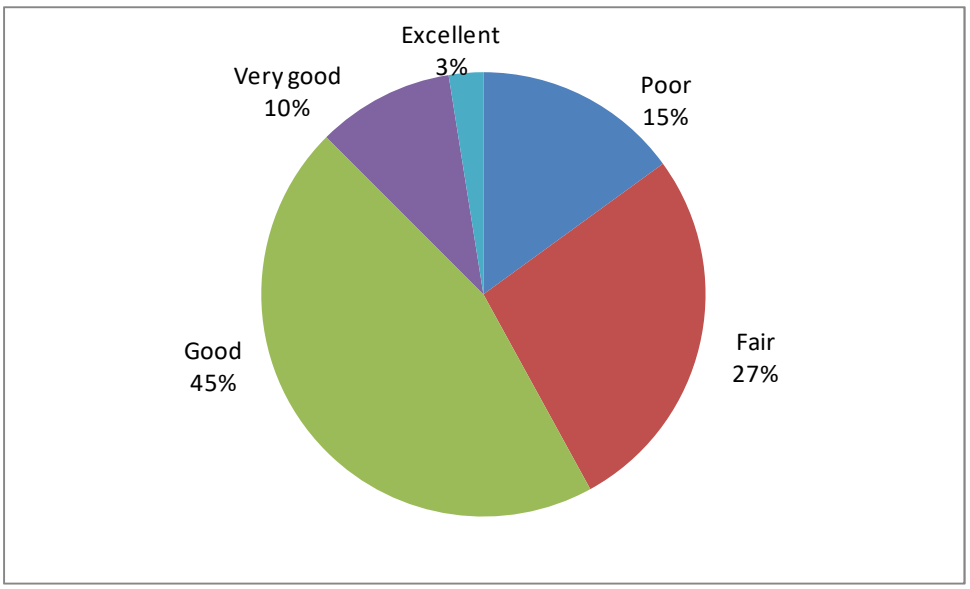
There are decreases in almost every state. However, there are many Americans without health insurance.

2.61



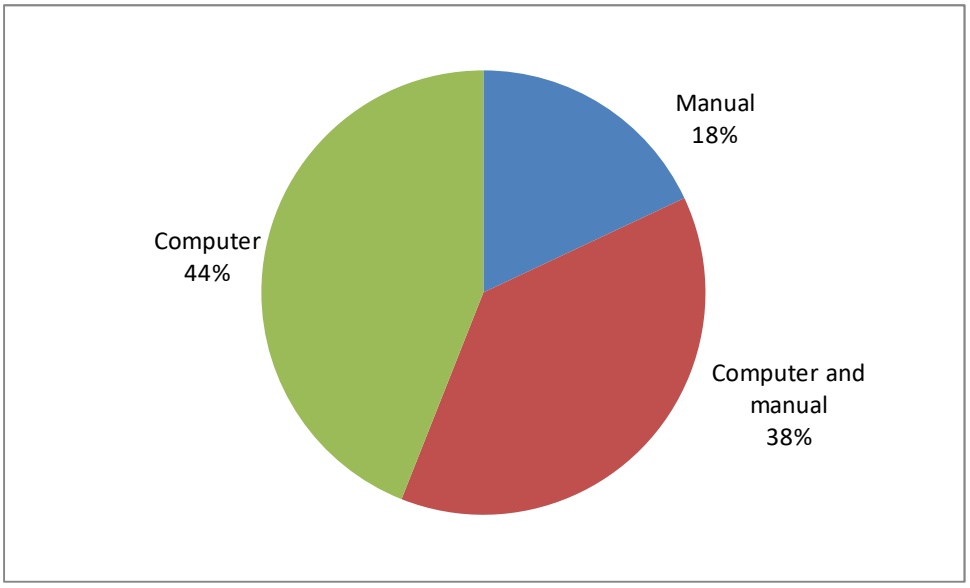
More students disagree than agree.

2.62



More than 40% rate the food as less than good.

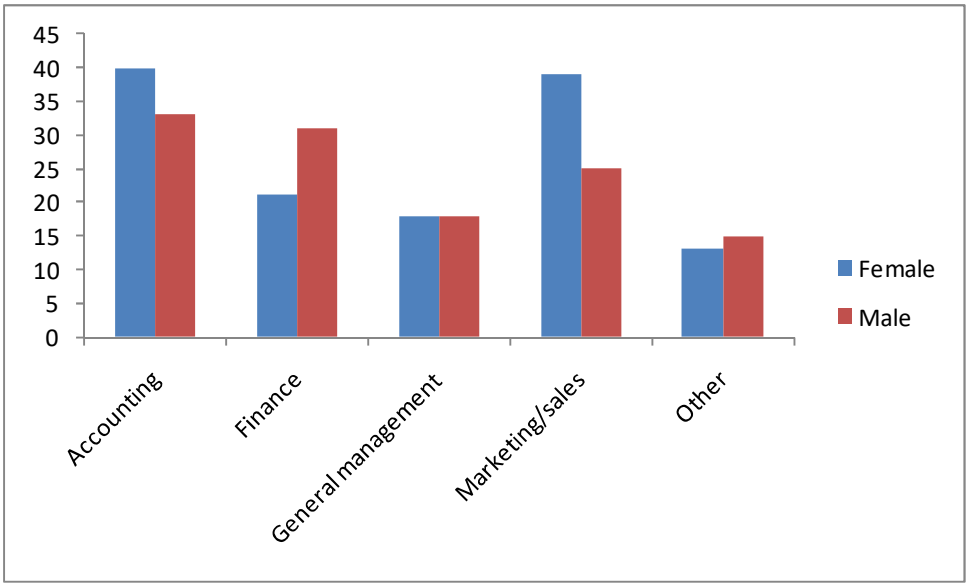
2.63



2.64

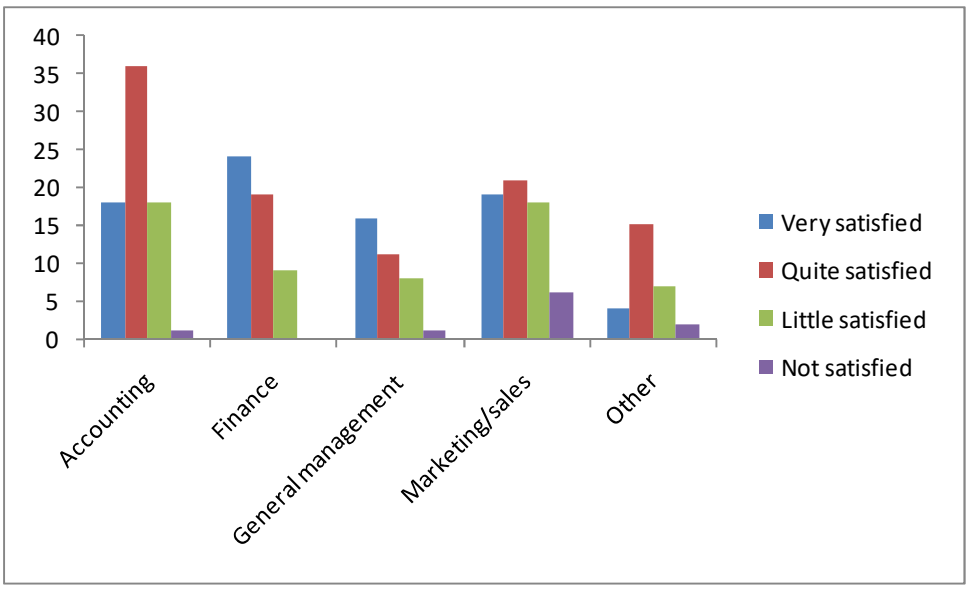
Customers with children rated the restaurant more highly than did customers with no children.

2.65



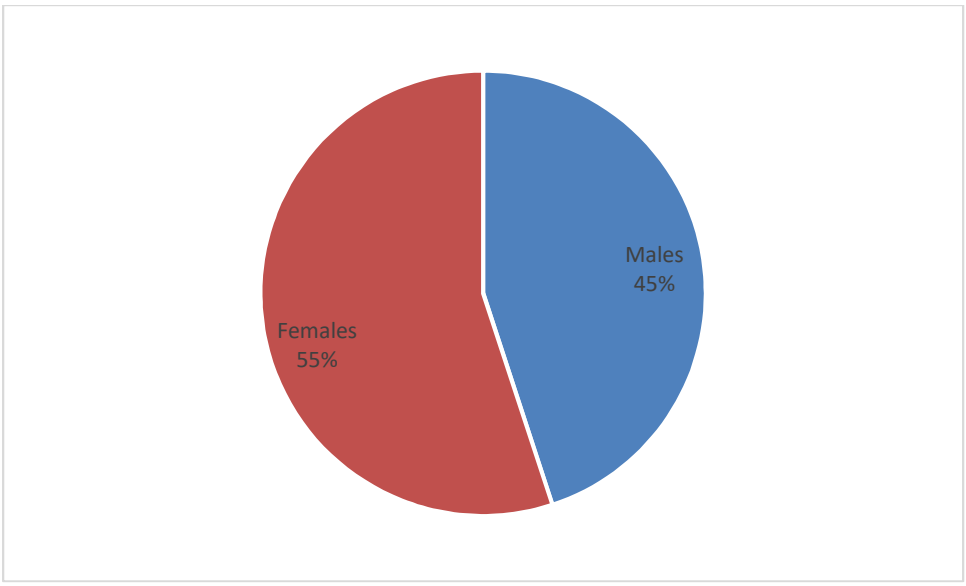
Males and females differ in their areas of employment. Females tend to choose accounting marketing/sales and males opt for finance.

b



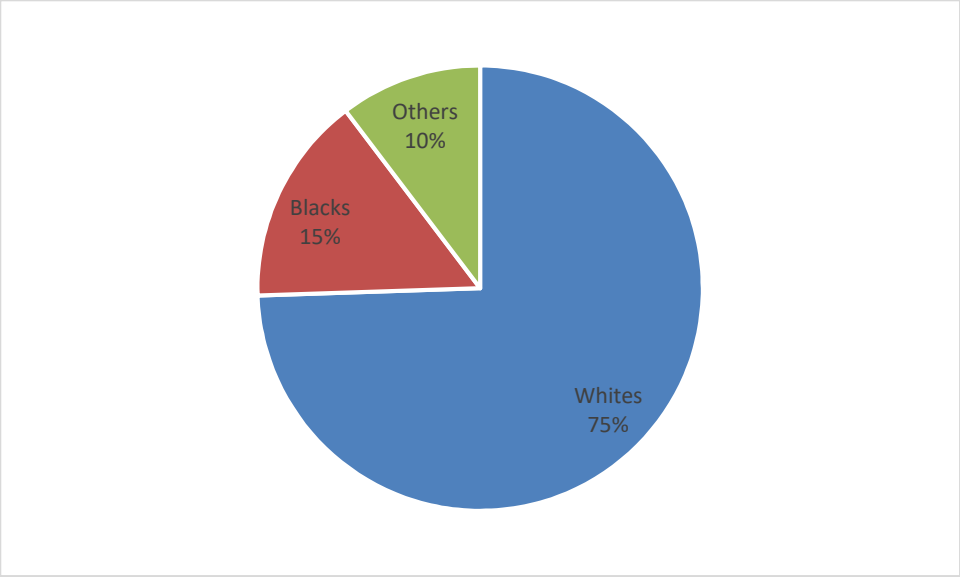
Area and job satisfaction are related. Graduates who work in finance and general management appear to be more satisfied than those in accounting, marketing/sales, and others.

2.66



The survey oversampled women slightly.

2.67

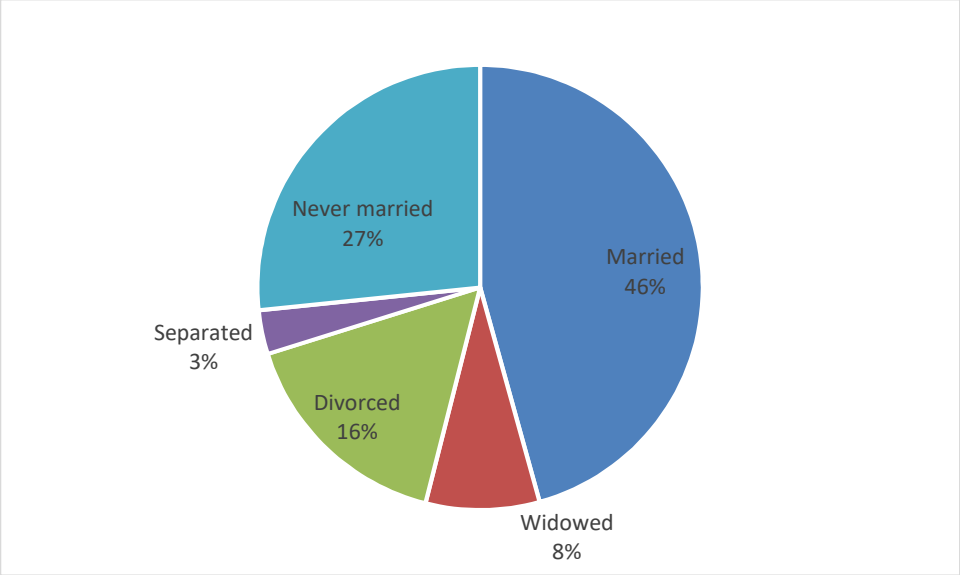


2.68a

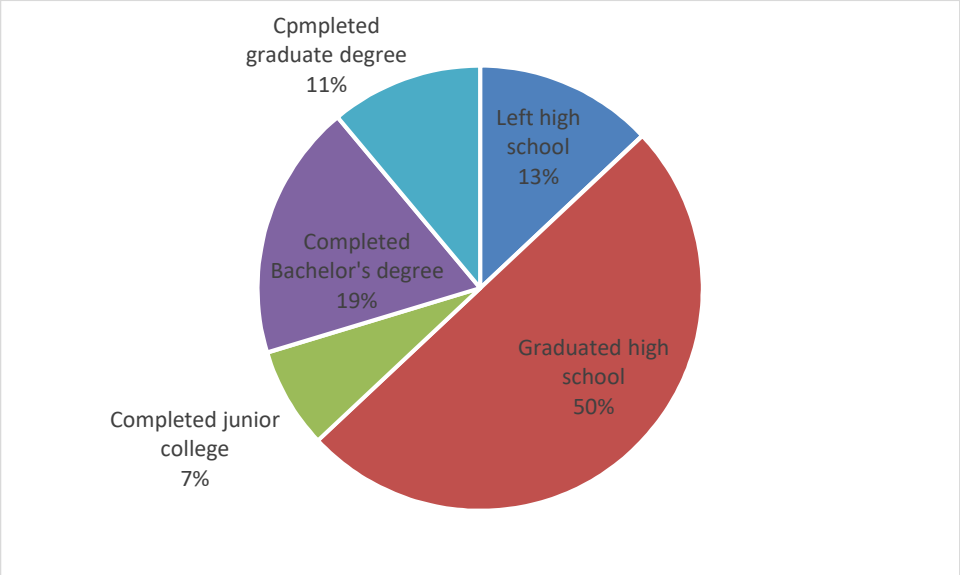
Married	1158
Widowed	209
Divorced	411
Separated	81
Never married	675

b. Pie chart

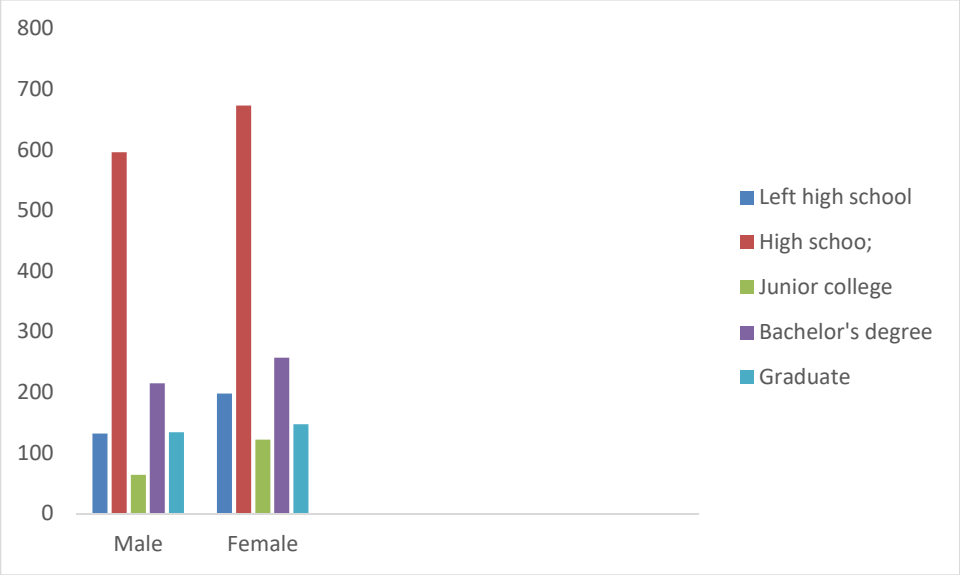
c.



2.69

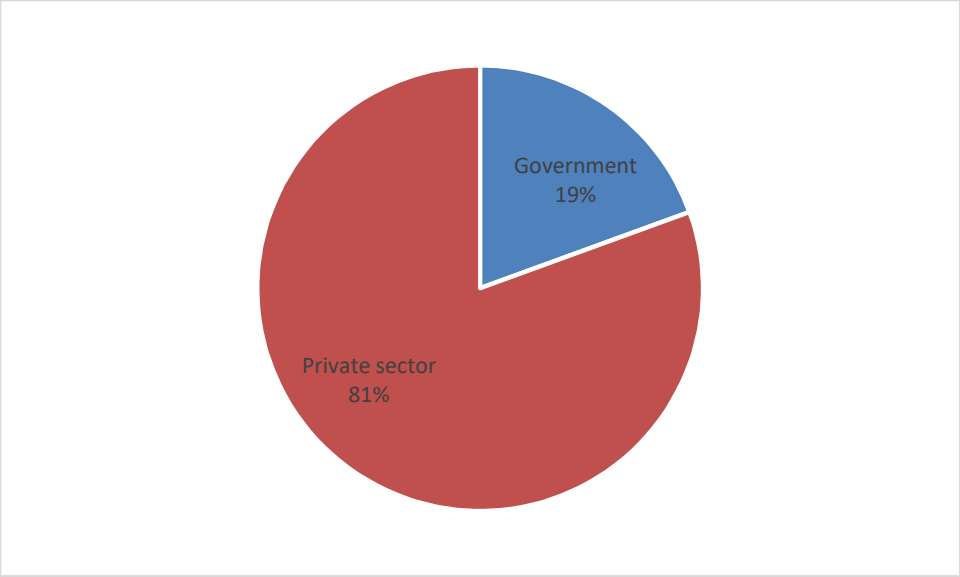


2.70

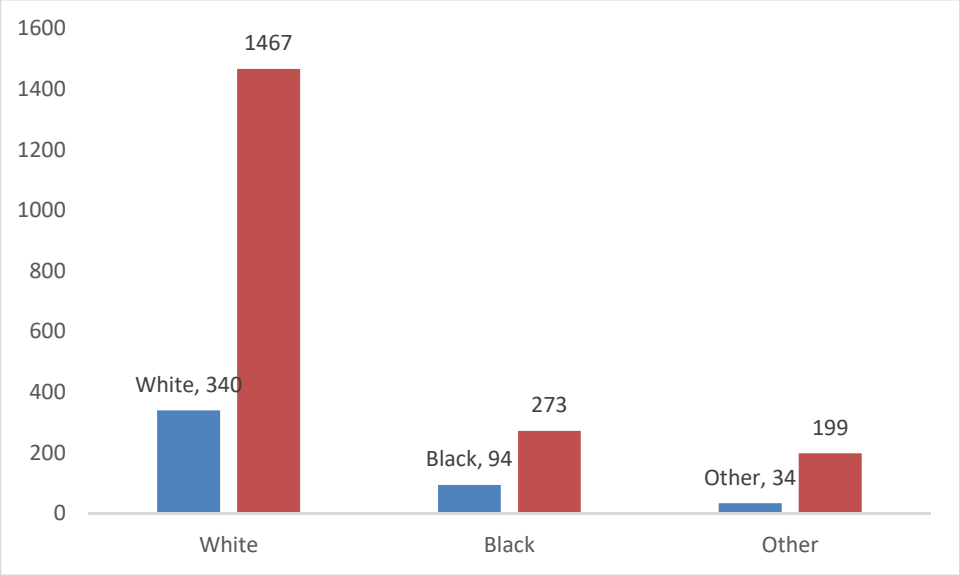


The patterns are similar.

2.71

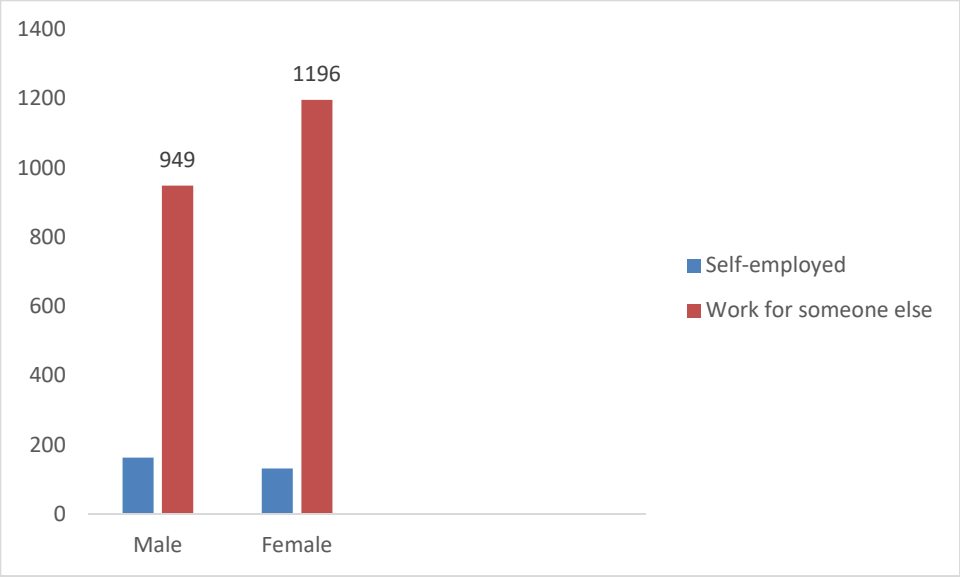


2.72



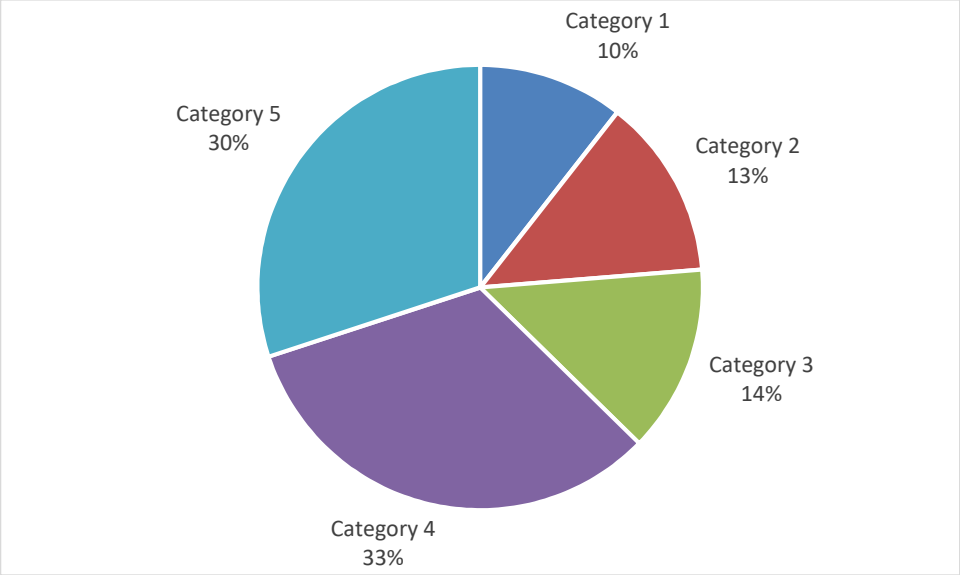
The patterns are similar.

2.73



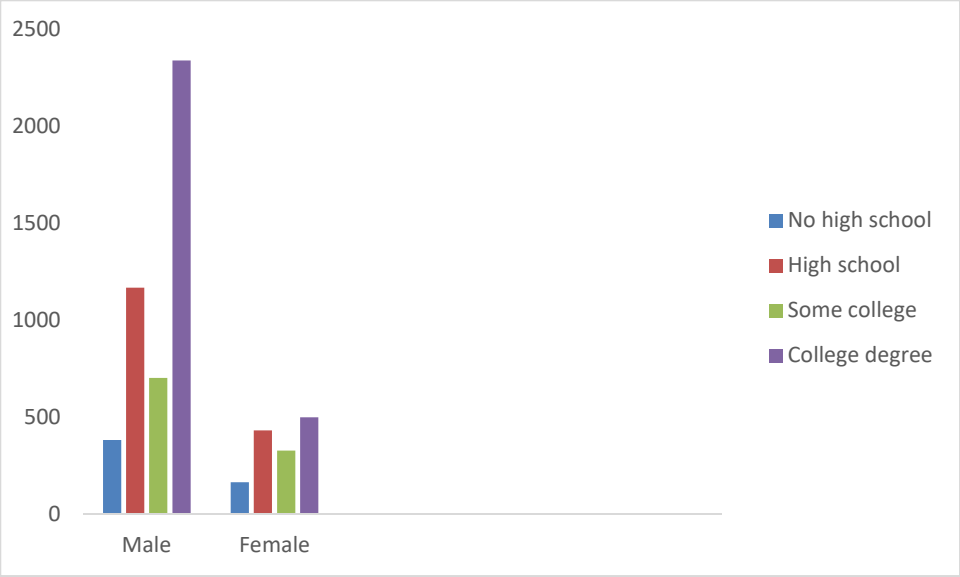
Males are slightly more likely to be self-employed than females.

2.74



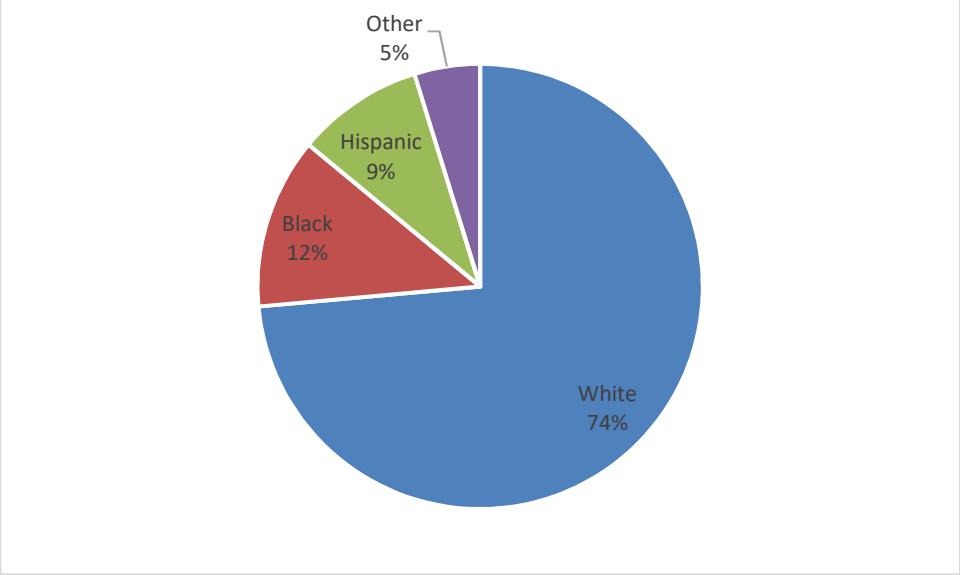
The ”married” categories (4 and 5) make up more than 60% of the households.

2.75



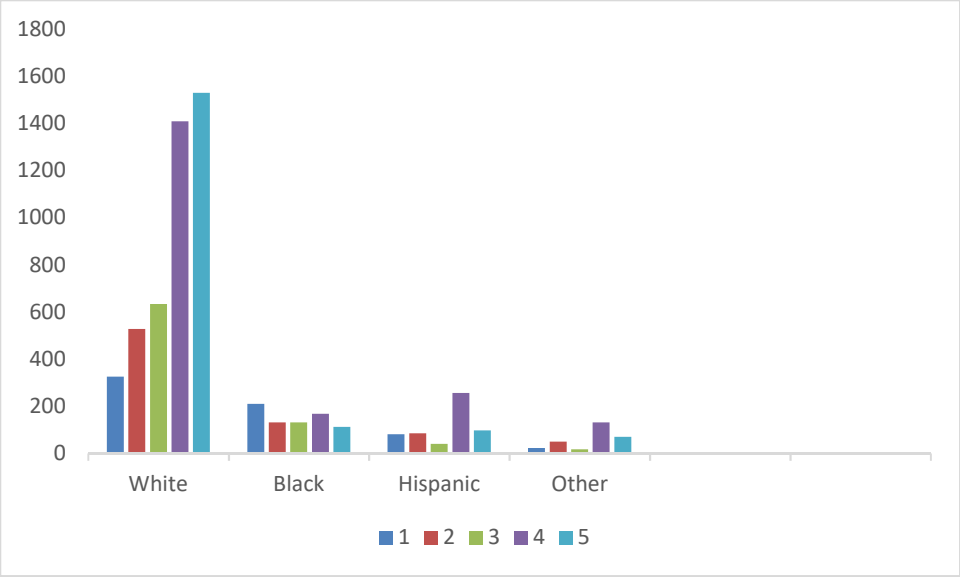
There are large differences between male and female heads of households.

2.76



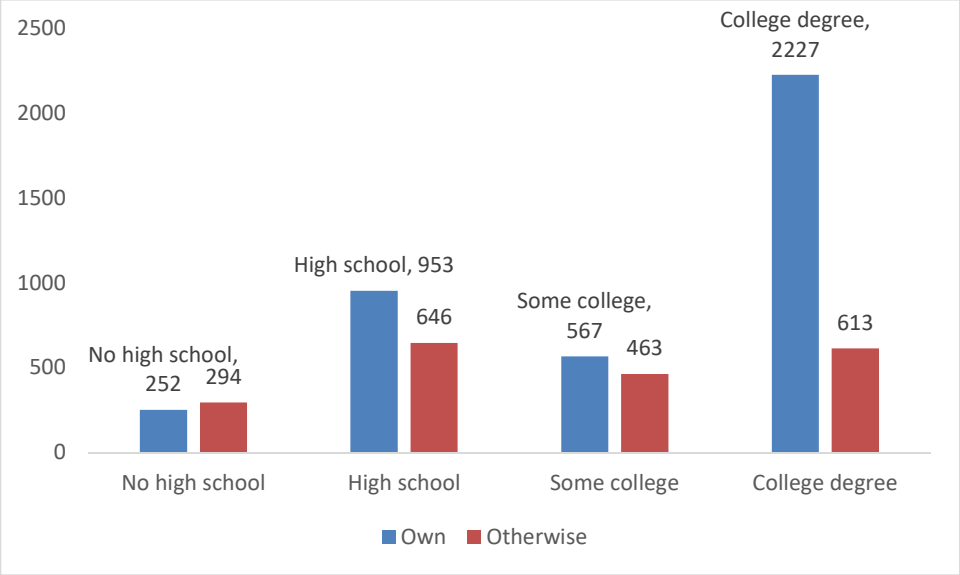
Whites make up three quarters of the survey.

2.77

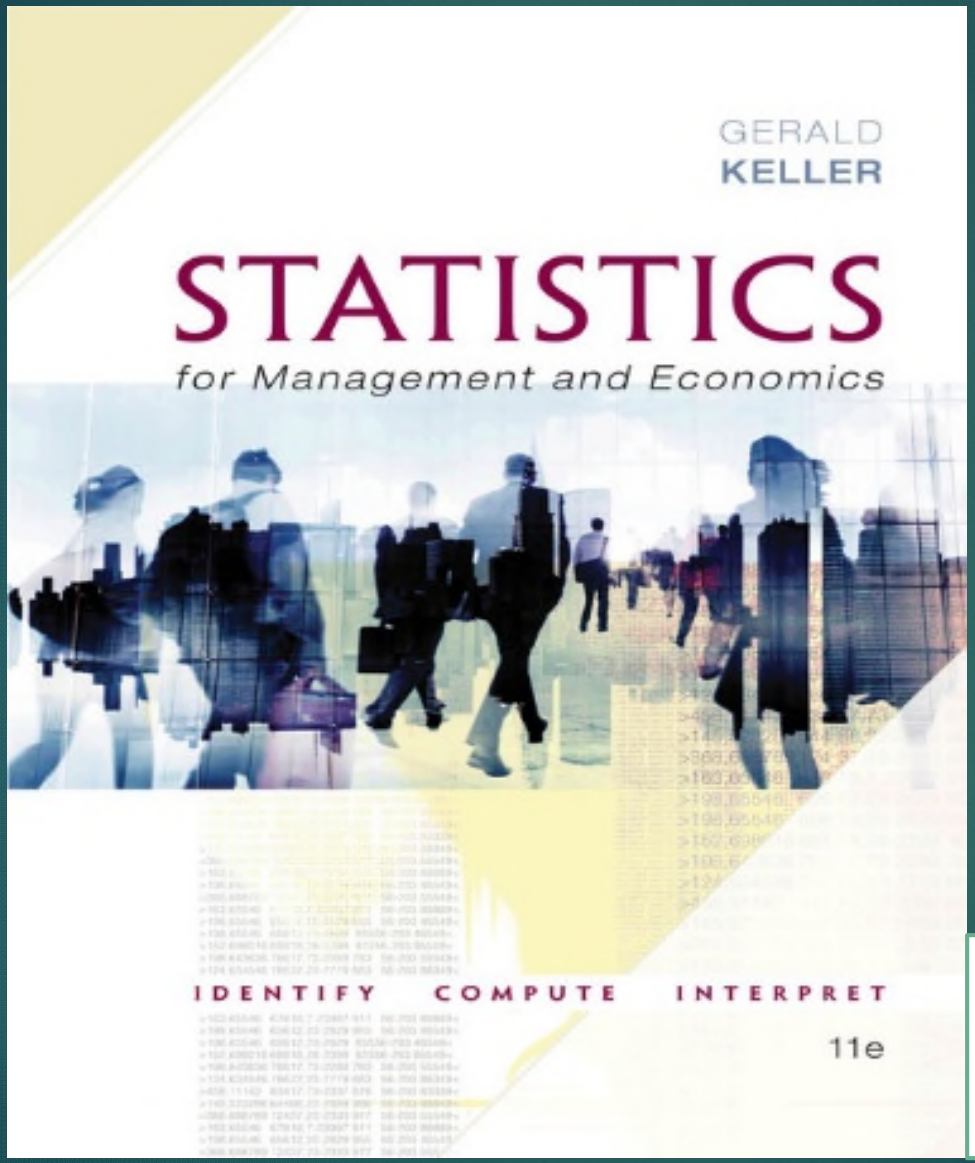


There are large differences between the four races in terms of family structure.

2.78



College degree holders are much more likely to own their homes.



Slides by:
Andrew Stephenson
Georgia Gwinnett College

Chapter Two

1.2

Graphical Descriptive Techniques 1

- Types of Data and Information
- Describing a Set of Nominal Data
- Describing the Relationship between Two Nominal Variables and Comparing Two or More Nominal Data Sets

Types of Data & Information

1.3

Definitions

A **variable** is some characteristic of a population or sample.

E.g. student grades.

Typically denoted with a capital letter: X, Y, Z…

The **values** of the variable are the range of possible values for a variable.

E.g. student marks (0..100)

Data are the ***observed values*** of a variable.

E.g. student marks: {67, 74, 71, 83, 93, 55, 48}

Types of Data & Information

1.4

Hierarchy of Data

Data (at least for purposes of Statistics) fall into three main groups:

Interval Data

Nominal Data

Ordinal Data

The data types can be placed in order of the permissible calculations. At the top of the list, we place the interval data type because virtually *all* computations are allowed. The nominal data type is at the bottom because *no* calculations other than determining frequencies are permitted.

Types of Data & Information

1.5

Interval data

- Real numbers, i.e. heights, weights, prices, etc.
- Also referred to as **quantitative** or **numerical**.

Arithmetic operations can be performed on Interval Data, thus its meaningful to talk about 2*Height, or Price + $1, and so on.

Types of Data & Information

Nominal Data

- The values of **nominal** data are ***categories.***

E.g. responses to questions about marital status, coded as:

Single = 1, Married = 2, Divorced = 3, Widowed = 4

These data are **categorical** in nature; arithmetic operations don't make any sense (e.g. does Widowed ÷ 2 = Married?!)

Nominal data are also called **qualitative** or **categorical**.

Types of Data & Information

Ordinal Data appear to be categorical in nature, but their values have an ***order***; a ranking to them:

E.g. College course rating system:

poor = 1, fair = 2, good = 3, very good = 4, excellent = 5

While its still not meaningful to do arithmetic on this data (e.g. does 2*fair = very good?!), we can say things like:

`excellent > poor` or **`fair < very good`**

That is, order is maintained no matter what numeric values are assigned to each category.

Types of Data & Information

As mentioned above,

- All calculations are permitted on **interval** data.
- Only calculations involving a ranking process are allowed for **ordinal** data.
- No calculations are allowed for **nominal** data, save counting the number of observations in each category.

This lends itself to the following "hierarchy of data"…

Types of Data & Information

Interval

Values are real numbers.

All calculations are valid.

Data may be treated as ordinal or nominal.

Ordinal

Values must represent the ranked order of the data.

Calculations based on an ordering process are valid.

Data may be treated as nominal but not as interval.

Nominal

Values are the arbitrary numbers that represent categories.

Only calculations based on the frequencies of occurrence are valid.

Data may not be treated as ordinal or interval.

Describing a Set of Nominal Data

Graphical & Tabular Techniques for Nominal Data…

The only allowable calculation on nominal data is to count the frequency of each value of the variable.

We can summarize the data in a table that presents the categories and their counts called a ***frequency distribution.***

A ***relative frequency distribution*** lists the categories and the proportion with which each occurs.

Example 2.1 Work Status in the GSS 2012 Survey

[GSS2012*] In Chapter 1 we briefly introduced the General Social Survey. In the 2012 survey respondents were asked the following question.

Last week were you working full time, part time, going to school, keeping house, or what? The responses were

1. Working full time
2. Working part time
3. Temporarily not working
4. Unemployed, laid off
5. Retired
6. School
7. Keeping house
8. Other

The responses were recorded using the codes 1, 2, 3, 4, 5, 6, 7, and 8, respectively.

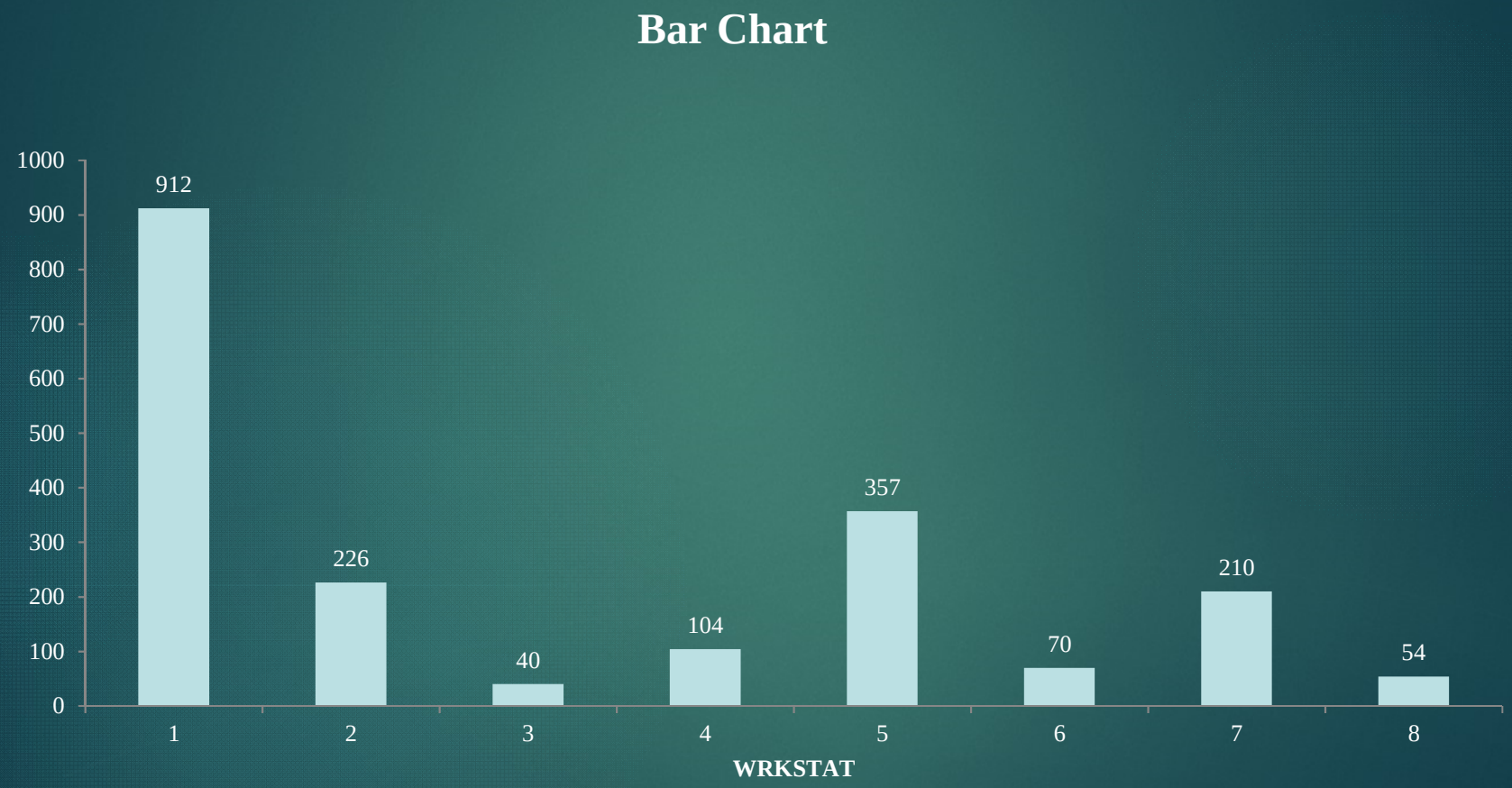
Describing a Set of Nominal Data

Frequency and Relative Frequency Distributions

Work Status	Code	Frequency	Relative Frequency (%)
Working full-time	1	912	46.2
Working part-time	2	226	11.5
Temporarily not working	3	40	2.0
Unemployed, laid off	4	104	5.3
Retired	5	357	18.1
School	6	70	3.5
Keeping house	7	210	10.6
Other	8	54	2.7

Describing a Set of Nominal Data

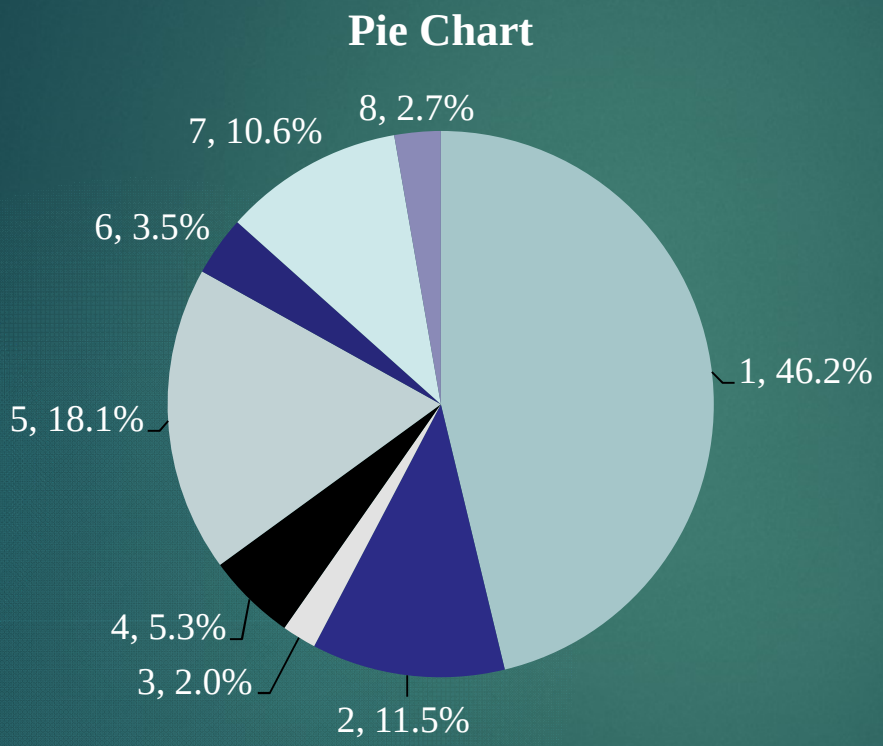
Nominal Data (Frequency)



Bar Charts are often used to display ***frequencies…***

Describing a Set of Nominal Data

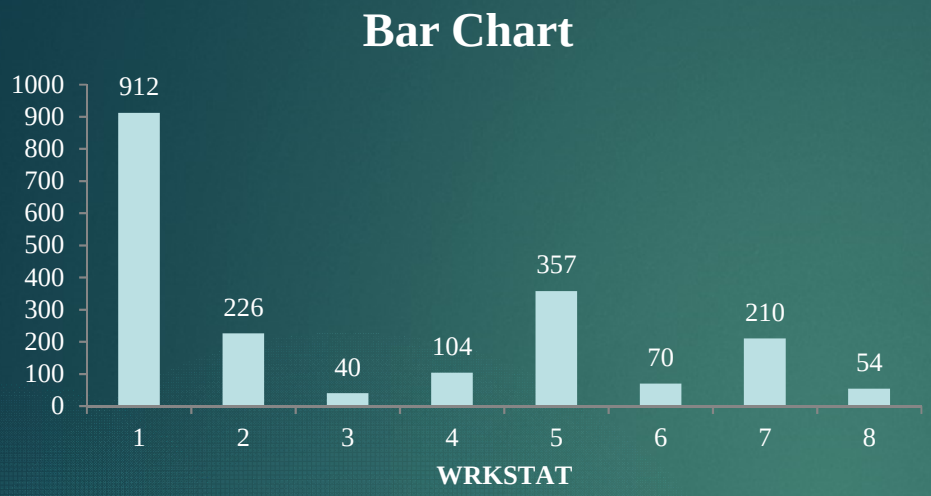
Nominal Data (Relative Frequency)



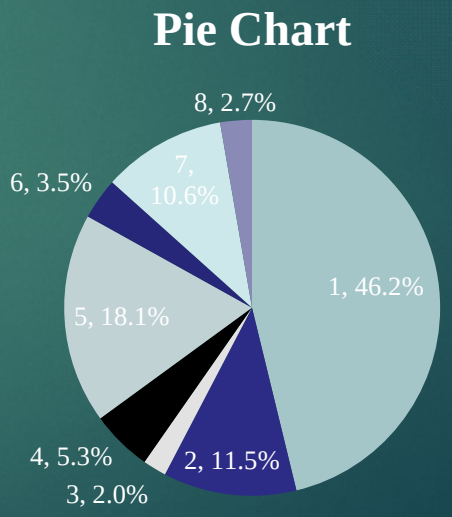
Pie Charts show ***relative frequencies…***

Describing a Set of Nominal Data

1.15



It is all the same ***information,*** (based on the same ***data***). Just different ***presentation.***



Describing the Relationship between Two Nominal Variables

To describe the relationship between two nominal variables, we must remember that we are permitted only to determine the frequency of the values. As a first step we need to produce a cross-classification table, which lists the frequency of each combination of the values of the two variables

Example 2.4 Newspaper Readership Survey

In a major North American city there are four competing newspapers: the *Globe and Mail (G&M), Post, Sun,* and *Star*. To help design advertising campaigns, the advertising managers of the newspapers need to know which segments of the newspaper market are reading their papers. A survey was conducted to analyze the relationship between newspapers read and occupation. A sample of newspaper readers was asked to report which newspaper they read: *Globe and Mail* (1) *Post* (2), *Star* (3), *Sun* (4), and to indicate whether they were blue-collar worker (1), white-collar worker (2), or professional (3). The responses are stored in Xm02-04 using the codes. Some of the data are listed here.

Example 2.4

Reader	Occupation	Newspaper
1	2	2
2	1	4
3	2	1
.	.	.
.	.	.
352	3	2
353	1	3
354	2	3

Determine whether the two nominal variables are related.

Describing the Relationship between Two Nominal Variables

Cross-Classification Table of Frequencies

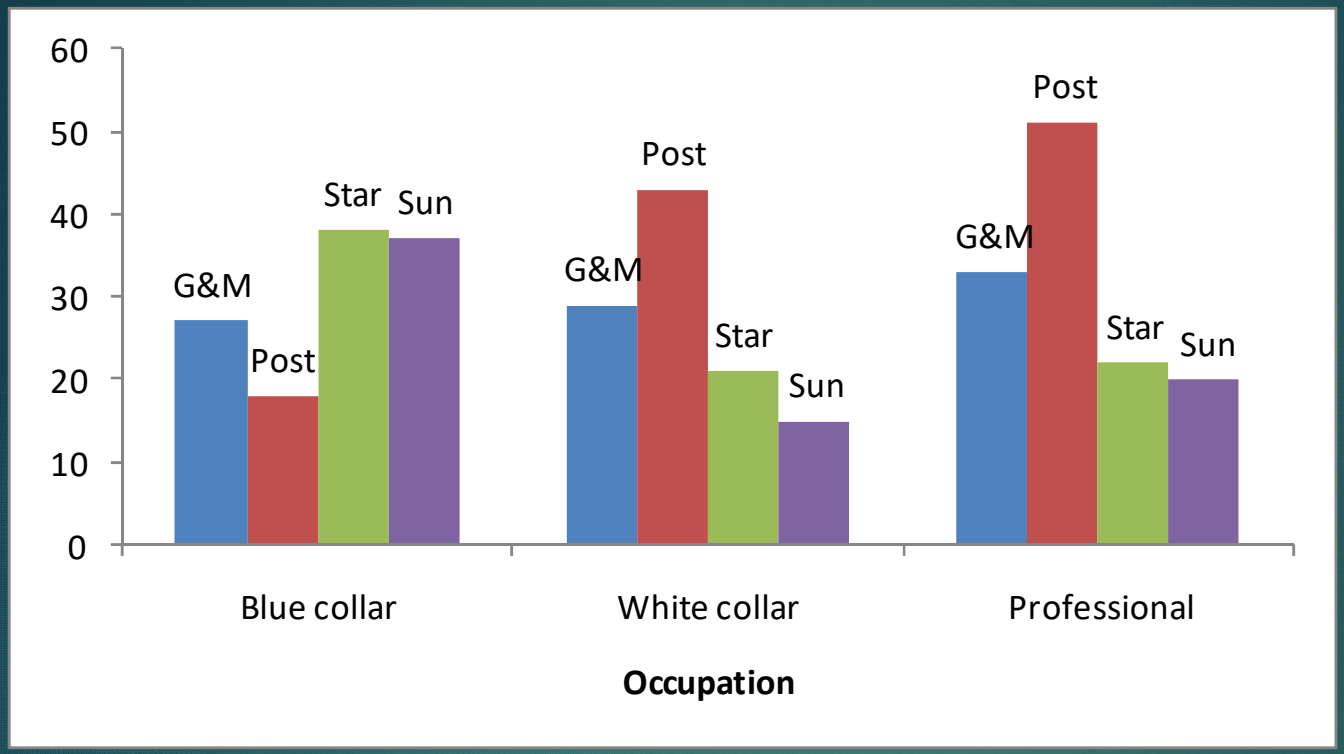
	Newspaper				
Occupation	*G&M*	*Post*	*Star*	*Sun*	Total
Blue collar	27	18	38	37	120
White collar	29	43	21	15	108
Professional	33	51	22	20	126
Total	89	112	81	72	354

Describing the Relationship between Two Nominal Variables

Row Relative Frequencies

	Newspaper				
Occupation	*G&M*	*Post*	*Star*	*Sun*	Total
Blue collar	.23	.15	.32	.31	1.00
White collar	.27	.40	.19	.14	1.00
Professional	.26	.40	.17	.16	1.00
Total	.25	.32	.23	.20	1.00

Graphing the Relationship between 2 Nominal Variables



The shapes of the bar charts for White-collar and Professional are very similar, but both differ considerably from Blue collar.

Graphing the Relationship between 2 Nominal Variables

INTERPRET

If the two variables are unrelated, the patterns exhibited in the bar charts should be approximately the same. If some relationship exists, then some bar charts will differ from others.

The graphs tell us the same story as did the table. The shapes of the bar charts for occupations 2 and 3 (White-collar and Professional) are very similar. Both differ considerably from the bar chart for occupation 1 (Blue-collar).