

1 Summarizing Categorical Data

Contingency tables:

Example 1: Smoking by Age Category

Age group	16-20	21-25	26-30	31-35
# smokers	35	28	26	19
# non-smokers	65	72	74	81

Bar Chart

What do you see?

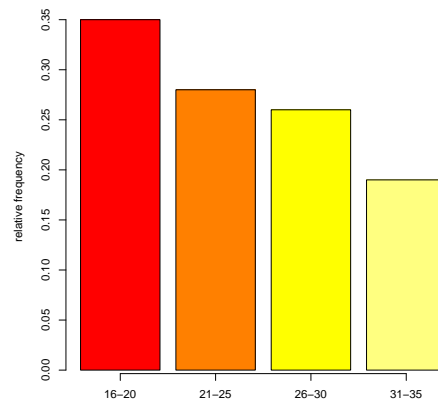


Figure 1: Relative frequency of smokers in each age group

If there were twice as many people in each part of the table, would the bar chart look any different?

Example 2: Employment and marital status

	Married	Widowed/Separated/Divorced	Never married
Employed	638	133	102
Unemployed	27	8	6
Not in labor force	35	12	20

Pie Charts

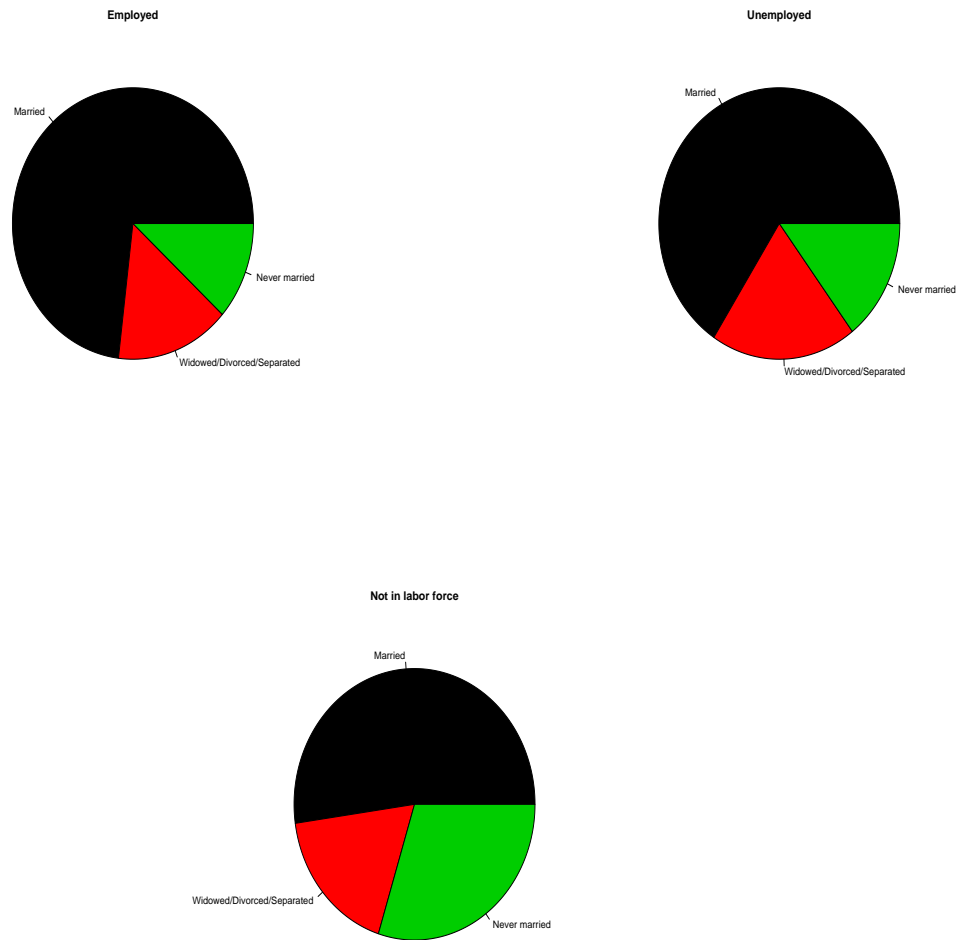


Figure 2: Top: Employed (left) and Unemployed (right), Bottom: Not in labor force

Bar Charts of Marital status in each Employment category

Which plot is the more informative? Counts of frequencies?

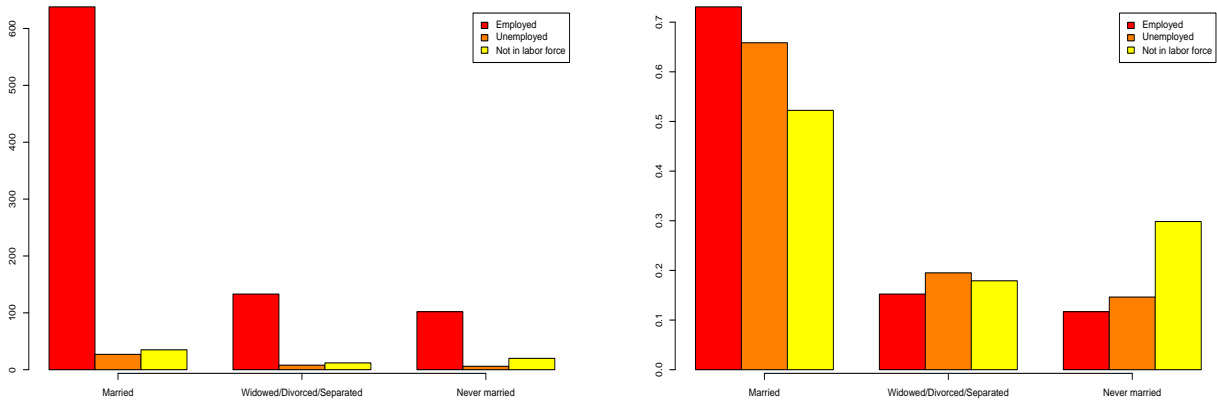


Figure 3: Left: bar plot w. counts. Right: bar plot with frequencies

Bar Charts of Employment status in each Marital category

Which plot is the more informative? Counts of frequencies?

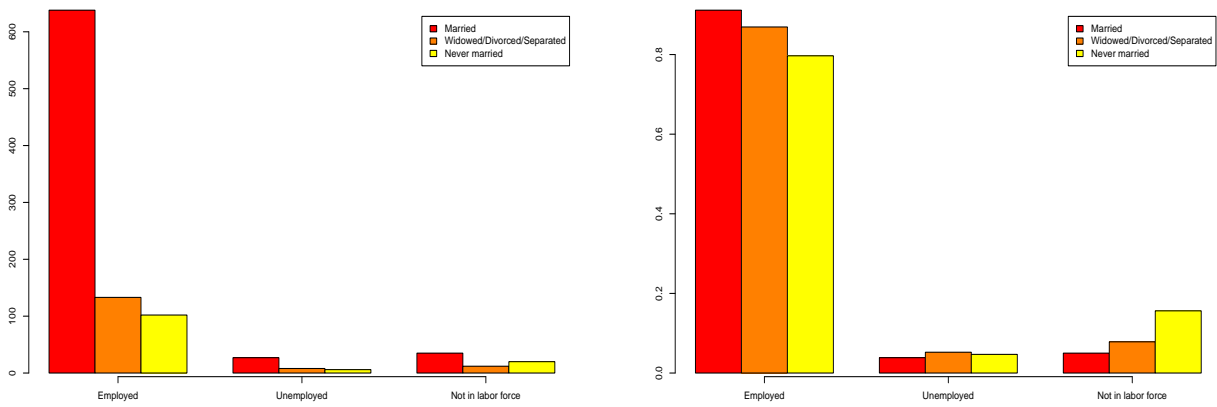
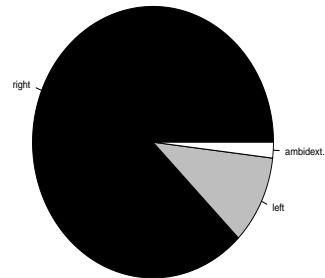


Figure 4: Left: bar plot w. counts. Right: bar plot with frequencies

Example 2: Handedness and Gender

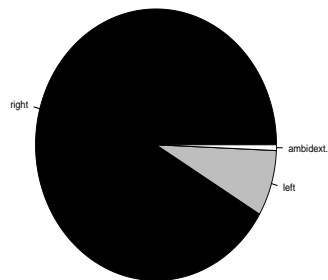
	Men	Women
# Right-handed	934	1070
# Left-handed	113	92
# Ambidextrous	20	8

Handedness - men



Pie Charts

Handedness - women



What do the pictures tell you about handedness? about gender differences?

Bar Charts Are gender differences easier to discern with bar charts?

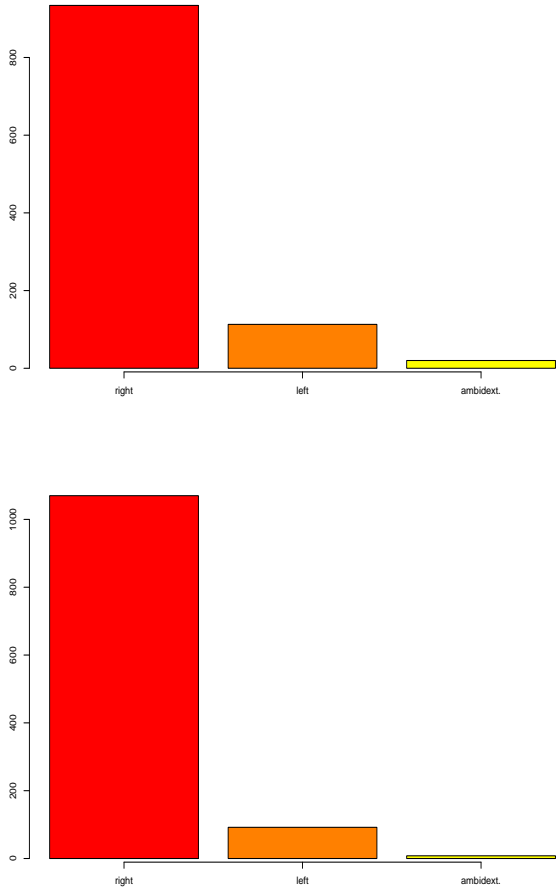
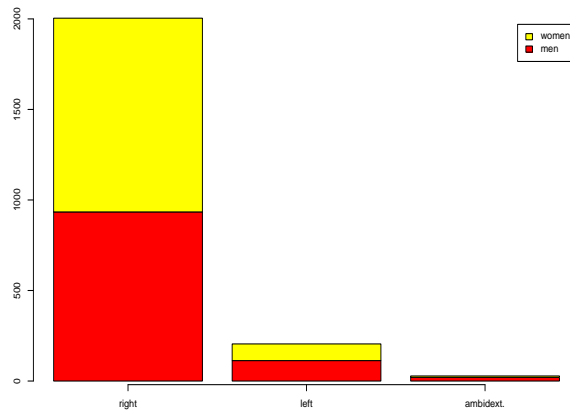
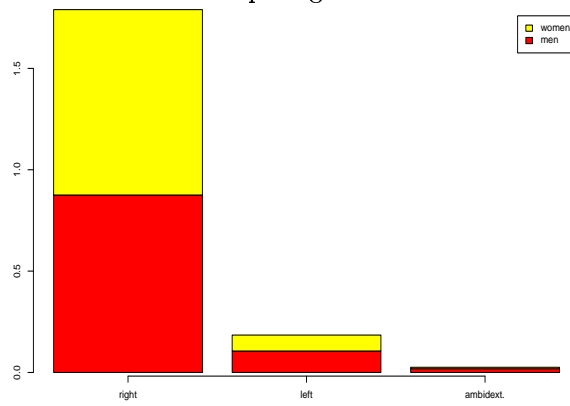


Figure 5: Top: Bar chart for men. Bottom: Bar chart for women.

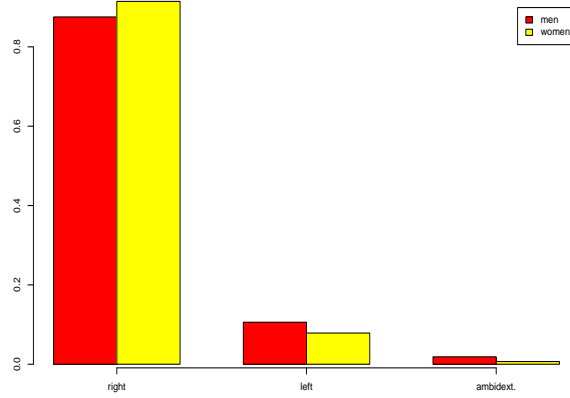
Bar Charts - Segmented and Side-by-Side Which bar chart is the most informative?



Comparing counts



Comparing relative frequencies



Comparing relative frequencies

2 Summarizing Quantitative Data

Stem and leaf plots:

```
                26 27 27 27 27   29 30 30 30 30   31 31 31 32 32
                33 33 33 33 33   34 34 34 35 35   36 36 36 37 37
                37 37 37 37 37   39 39 39 39 39   39 39 40 41 42
                42 42 42 42 43   43 43 43 43 43   43 43 44 44 44
                44 44 44 45 45   45 45 45 45 45   46 46 46 46 46
                46 47 47 47 47   47 47 48 48 48   48 48 48 48 48
                49 49 49 49 50   50 51 51 51 51   51 52 52 52 52
Exam test scores: 52 53 53 53 53   53 54 54 54 54   54 55 55 55 55
                56 56 56 56 57   57 57 57 58 58   58 58 58 58 58
                58 59 59 59 59   60 60 60 60 60   60 61 61 61 61
                61 61 62 62 62   63 63 64 65 66   66 66 67 67 67
                67 68 68 69 69   69 69 69 69 69   69 71 71 72 73
                74 74 74 75 75   76 76 78 80 80   80 80 81 81 81
                82 82 83 83 83   83 84 84 84 84   84 84 84 90 90
                90 91 91 91 92   92 92 93 93 93   93 95 95
```

223 tests, scores ranging from 26 to 95.

```
2 | 677779
3 | 00001112233333444
3 | 5566677777779999999
4 | 0122223333333444444
4 | 5555555666666777778888889999
5 | 00111112222333344444
5 | 5556666677778888889999
6 | 00000011111222334
6 | 56667778899999999
7 | 1123444
7 | 55668
8 | 00001112233334444444
8 |
9 | 0001112223333
9 | 55
```

A Stem-and-leaf plot preserves all the data, tells you about the shape, center and spread of the data.

Histograms

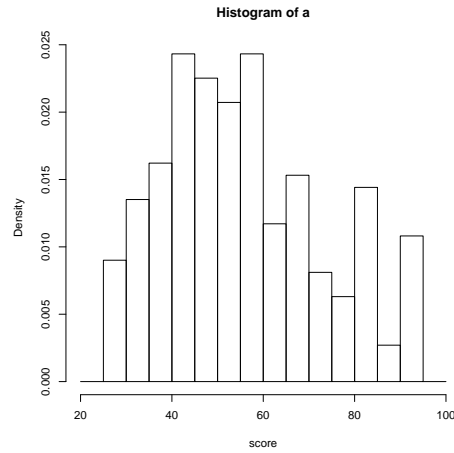


Figure 6: Histogram of scores, bar-width = 5 units

What to put on the axes? The area of each bar (times 100 %) should equal the relative frequency (in %).

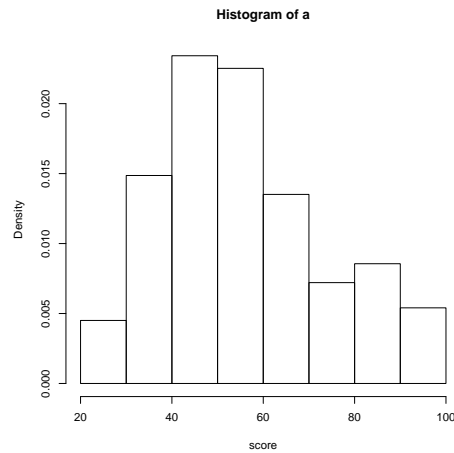


Figure 7: Histogram of scores, bar-width = 10 units

You can vary the width of the bins in the histogram.

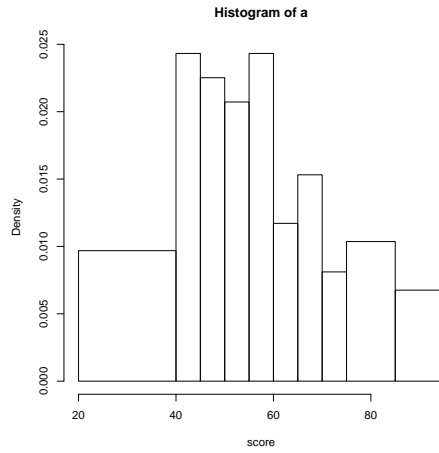


Figure 8: Histogram of scores, bar-width are mixed

You can even mix the widths of the bins in the same histogram.

More on histograms

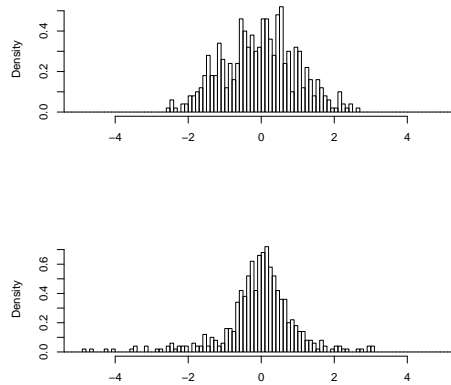


Figure 9: Tails of a distribution. Same center and spread, but short vs long tails.

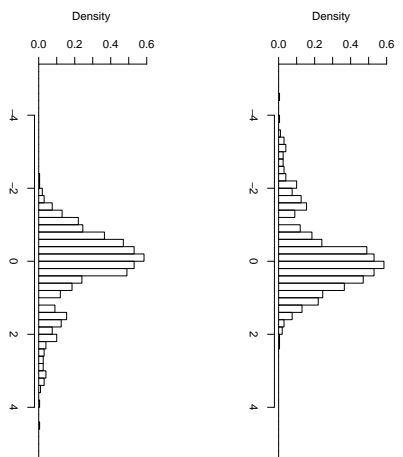


Figure 10: Tails of a distribution. Long left and right tails.

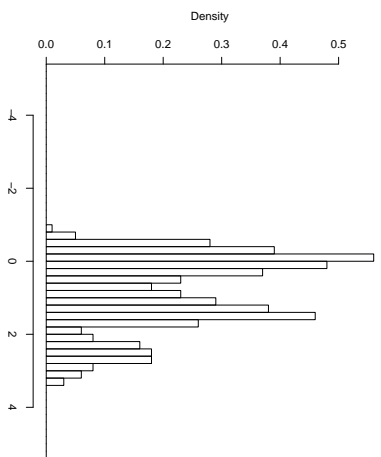


Figure 11: Multi-modality

Box-and-Whiskers. Comparing Distributions

Test scores on a Monday and on a Friday. What differences do you see?

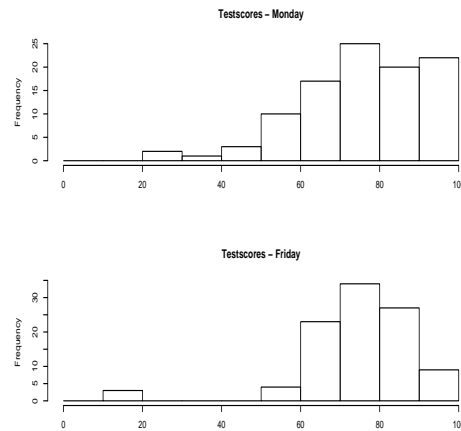


Figure 12: Histograms of Monday (left) and Friday (right) test scores

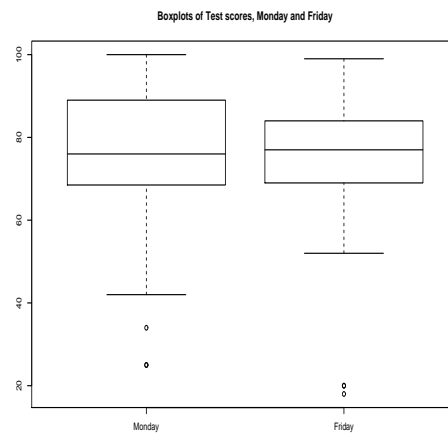


Figure 13: Box plots of Monday (top) and Friday (bottom) test scores