



Statistik I — bis jetzt

K 1 Einführung

K2 Beschreibende Statistik

- 2.1 Fragen
- 2.2 Beschreibung eines Datensatzes
- 2.3 Datentypen, Daten Schwierigkeiten
- 2.4 Schwierigkeiten mit Daten
- 2.5 Beschreibung kategorialer Variablen
- 2.6 Beschreibung von stetigen Variablen
- 2.7 Statistiken / Kenngrößen

K3 Statistische Graphik

- 3.1 Graphische Darstellungen stetiger Variablen
(Punktplots („Dotplots“), Histogramme, Boxplots)

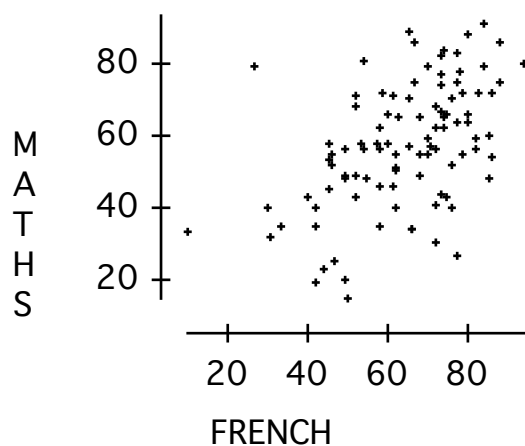
und heute

- 3.2 Bivariat stetig — Streudiagramme
- 3.3 Kategoriale Variablen
Säulendiagramme, Spineplots, Kreisdiagramme
- 3.4 Zeitreihen

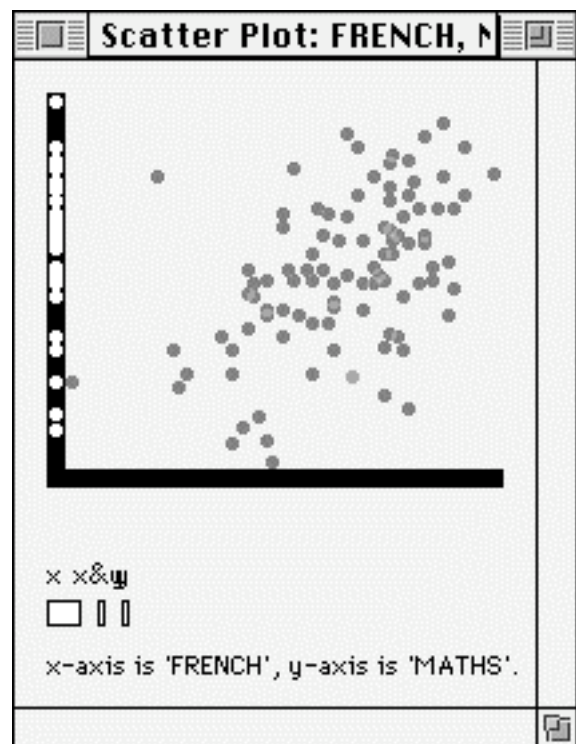
3.2 Streudiagramme

Darstellungen von zwei stetigen Variablen,

Eine Variable wird als Y (vertikale Achse) und eine als X (horizontale Achse) genommen:



Data Desk



MANET

Formatierungsparameter:

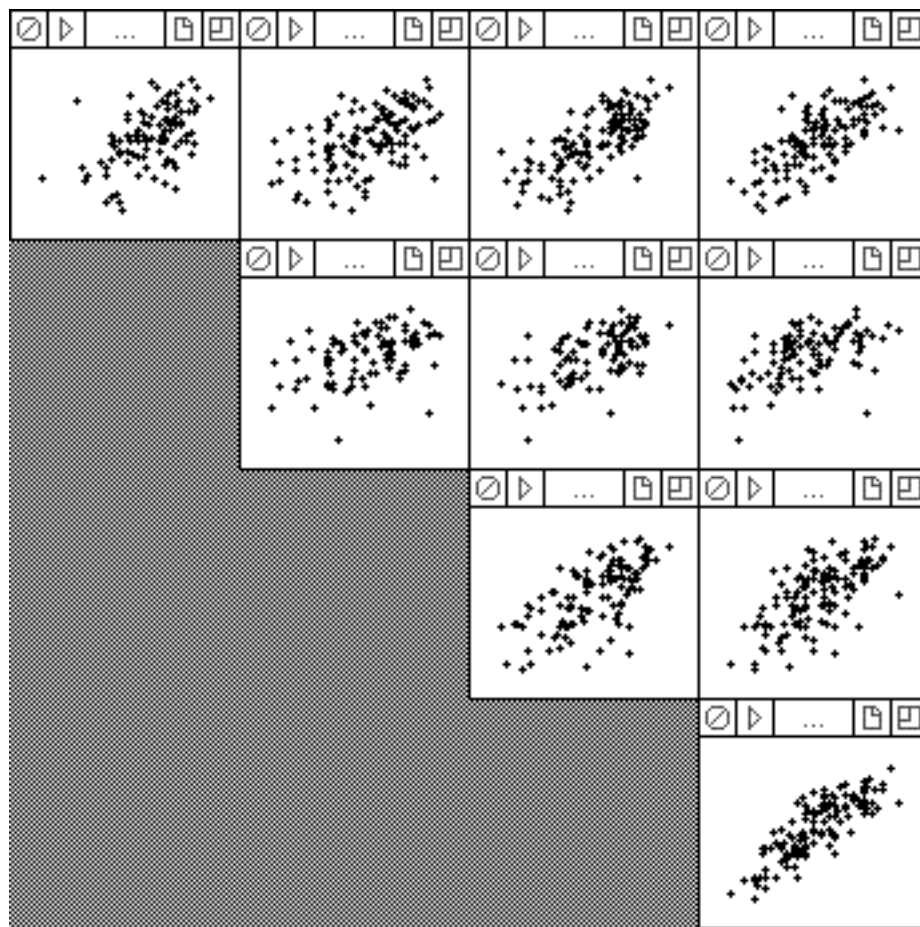
Punkt-Größe und -Farbe

Skalierung, Aspekt

Beschriftung, Legende

Für mehrere stetige Variablen kann man eine Matrix von Streudiagrammen zeichnen:

(Irische Prüfungsergebnisse, Mathematik, Französisch, Geschichte, Geographie, Naturwissenschaft)



Das Diagramm an der Stelle (i,j) ist das Streudiagramm von X_i gegen X_j ($i \neq j$). Auf der Diagonale findet man manchmal die Namen der Variablen, manchmal Histogramme oder andere univariate Darstellungen.

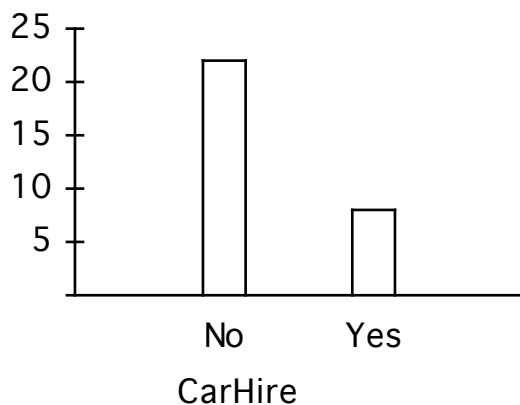
3.3 Darstellungen kategorialer Daten

3.3.1 (a) Säulendiagramme

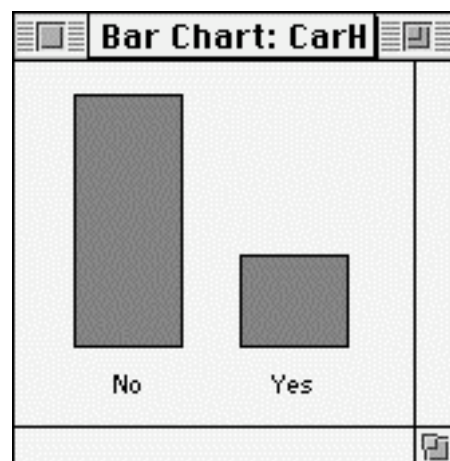
Es gibt eine Säule für jede Kategorie.

Alle Säulen sind gleich breit.

Die Säulenhöhe stellt die Anzahl der Fälle in der Kategorie dar.



Data Desk



MANET

Eigenschaften:

Größe

Skalierung und Abstand zwischen den Säulen

Beschriftung und Legende

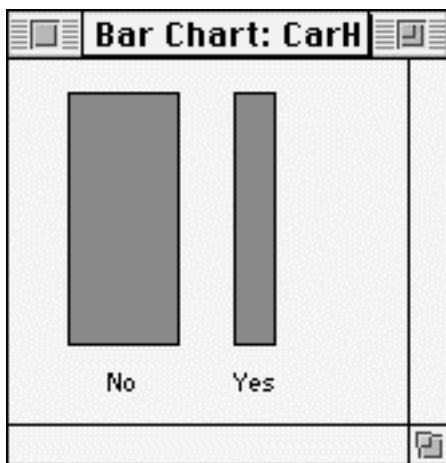
Farbe und Schattierung

3.3.1 (b) Spine Plots

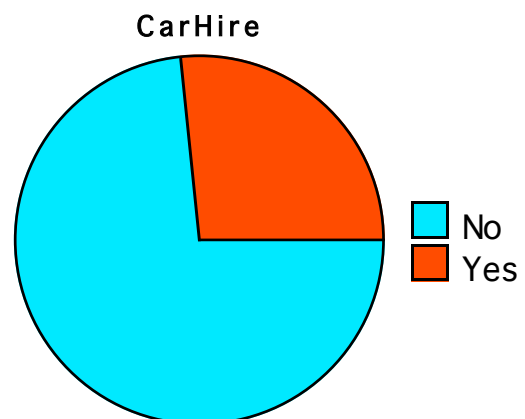
Es gibt eine Säule für jede Kategorie.

Alle Säulen sind gleich hoch.

Die Säulenbreite stellt die Anzahl der Fälle in der Kategorie dar.



MANET



Data Desk

(c) Kreisdiagramme (Tortendiagramme)

Der Kreis wird in Flächen zerlegt.

Jedes Stück ist der Anzahl der Fälle proportional.



Vergleiche

(a) Säulendiagramme

Am besten für direkte Vergleiche der Kategorien

Höhe $\propto n_i$

(b) Spine Plots

Zum Vergleichen von selektierten Proportionen

Breite $\propto n_i$

(c) Kreisdiagramme

Betont Proportionen als Anteile eines Ganzen

Winkel $\propto n_i$



3.3.2 Zwei kategorialen Variablen

z.B. Autovermietung und Besuch nur in Irland

Eine Kontingenztafel

	Rows are levels of Columns are levels of No Selector		
	No	Only	CarHire IrelandOnly1
No	12	10	
Yes	5	3	
table contents: Count			

Im allgemeinen gibt es r Zeilen (engl. „rows“), c Spalten (engl. „columns“) und $r \times c$ Zellen insgesamt.

Darstellungsvariante:

Die Daten können als Zahlen, Zeilenprozent, Spaltenprozent oder, seltener, als Gesamtprozent angegeben werden.

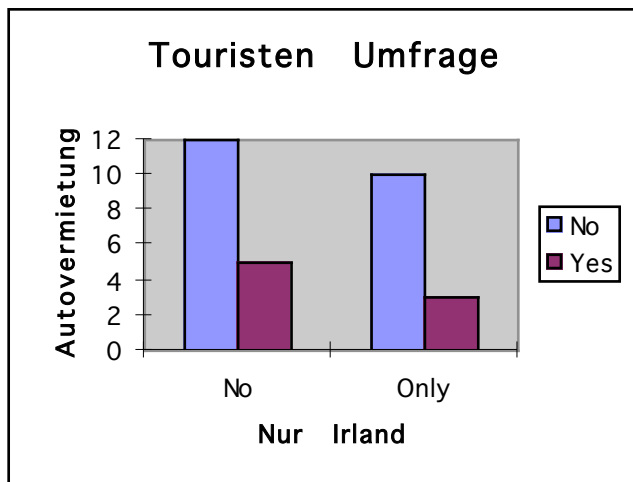
Die Reihenfolge der Zeilen bzw. Spalten kann geändert werden.

Die Zeilenvariable und die Spaltenvariable können vertauscht werden.



Säulendiagramme für zwei kategorielle Variablen

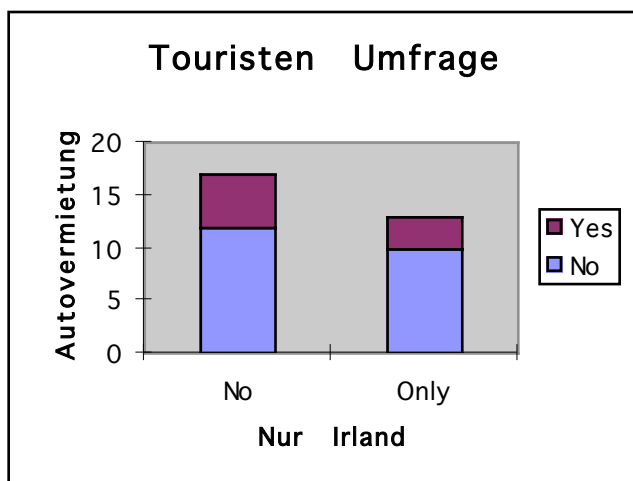
(a) Nebeneinander



Excel

Die vier Kombinationen sind direkt vergleichbar.

(b) Gestapelt (nach absoluten Größen)

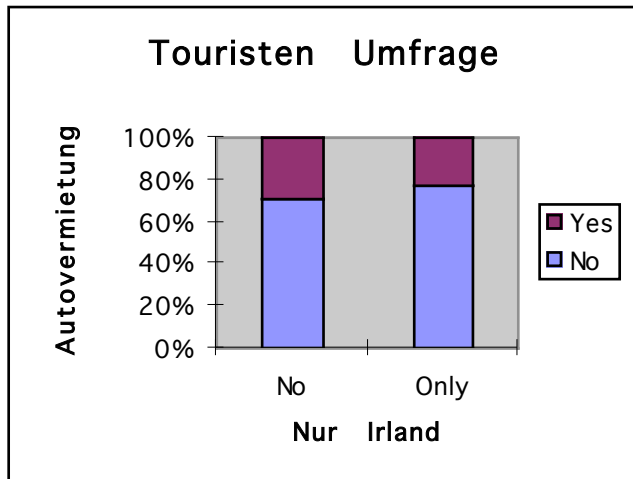


Excel

Proportionen werden betont.

Bei ungleich großen Gruppen ist es schwierig, die Proportionen innerhalb der Gruppen zu vergleichen.

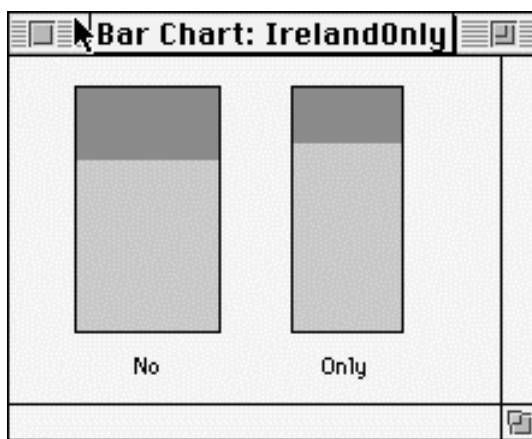
(c) gestapelt (nach Prozent)



Excel

Die Proportionen sind jetzt direkt vergleichbar, nur sind die Gruppengrößen vernachlässigt.

(d) gestapelt (nach Breite): Spineplots



MANET

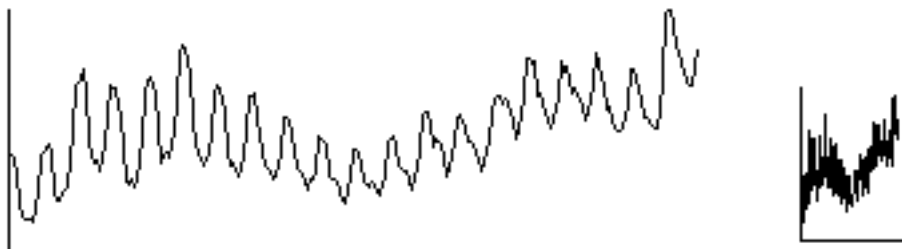
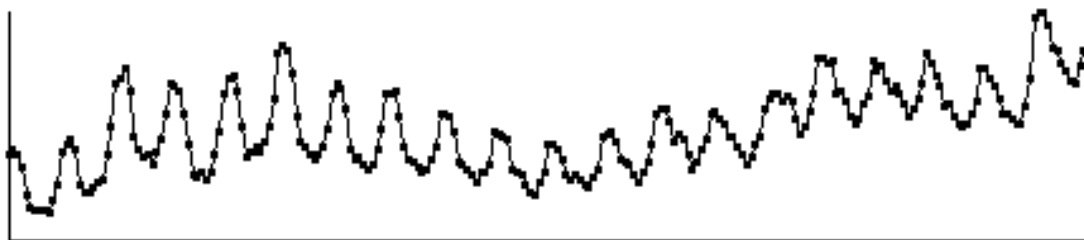
Hier sieht man den Vorteil des Spineplots: Proportionen vergleichen zu können, ohne die Gruppengrößen zu vernachlässigen.

3.4 Zeitreihen

Zeitabhängige Beobachtungen bilden eine Zeitreihe $\{x_t\}$

z.B. Temperatur, Aktienpreise, Verkaufszahlen

Die Werte werden wie in einem Streudiagramm dargestellt, mit Zeit immer auf der horizontalen Achse.



Fünf Abbildungen derselben monatlichen Arbeitslosenzahlen aus Kanada von 1952 bis 1971.



Bei einer Zeitreihengraphik sind u.a. wichtig:

Bildgröße

Skalierung

Ursprung

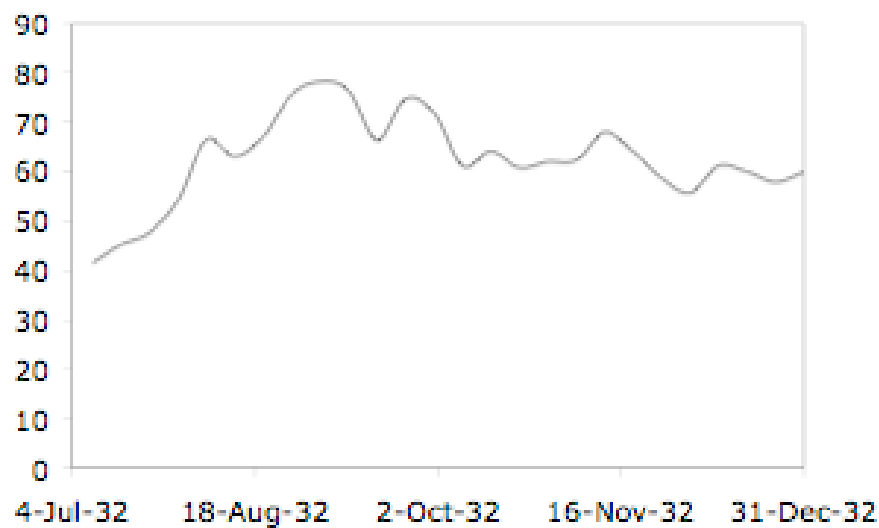
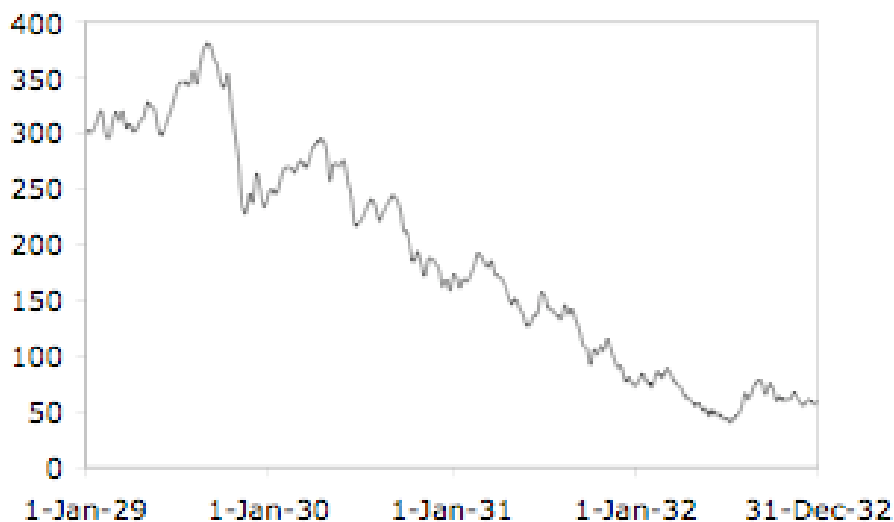
Untere und obere Grenzen

Zeitbeschriftungen

Aspekt ratio (Seitenverhältnis)

Darstellung (Punkte, Linien, Glättungen)

z.B. Dow Jones 1929-1932 und Juli bis Dez 1932





Statistik und die Nachrichten — AZ 3.5.06

s.2 AIDS

Nach Schätzung des Robert-Koch-Instituts leben in der Bundesrepublik 49000 Frauen und Männer mit dem Krankheitserreger HIV im Leib.

In Südafrika gibt es 5,3 Millionen Menschen, die das Virus in sich tragen. Das entspricht 21,5 Prozent der Bevölkerung.

s.9 Reichensteuer

Die jetzige Regelung dürfte den öffentlichen Haushalten lediglich bis zu 300 Millionen Euro Mehreinnahmen bringen.

Wie werden solche Zahlen geschätzt?