



A Gentle Excursion into

Data Visualization

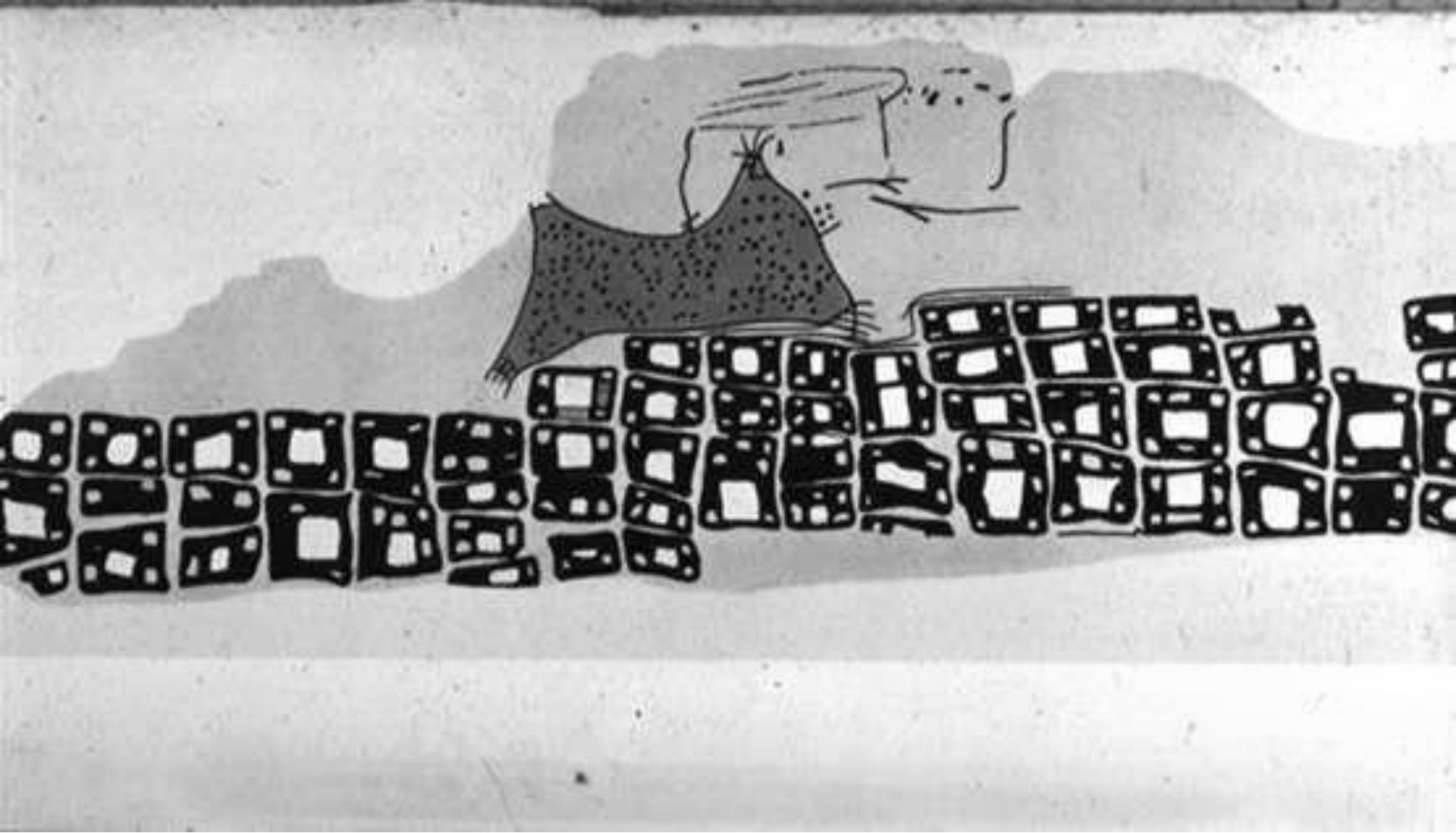
Pavel Vaněček

1. Selected moments from history
2. Univariate charts
3. Bivariate charts
4. Multivariate charts
5. Inspiration
6. Live demo

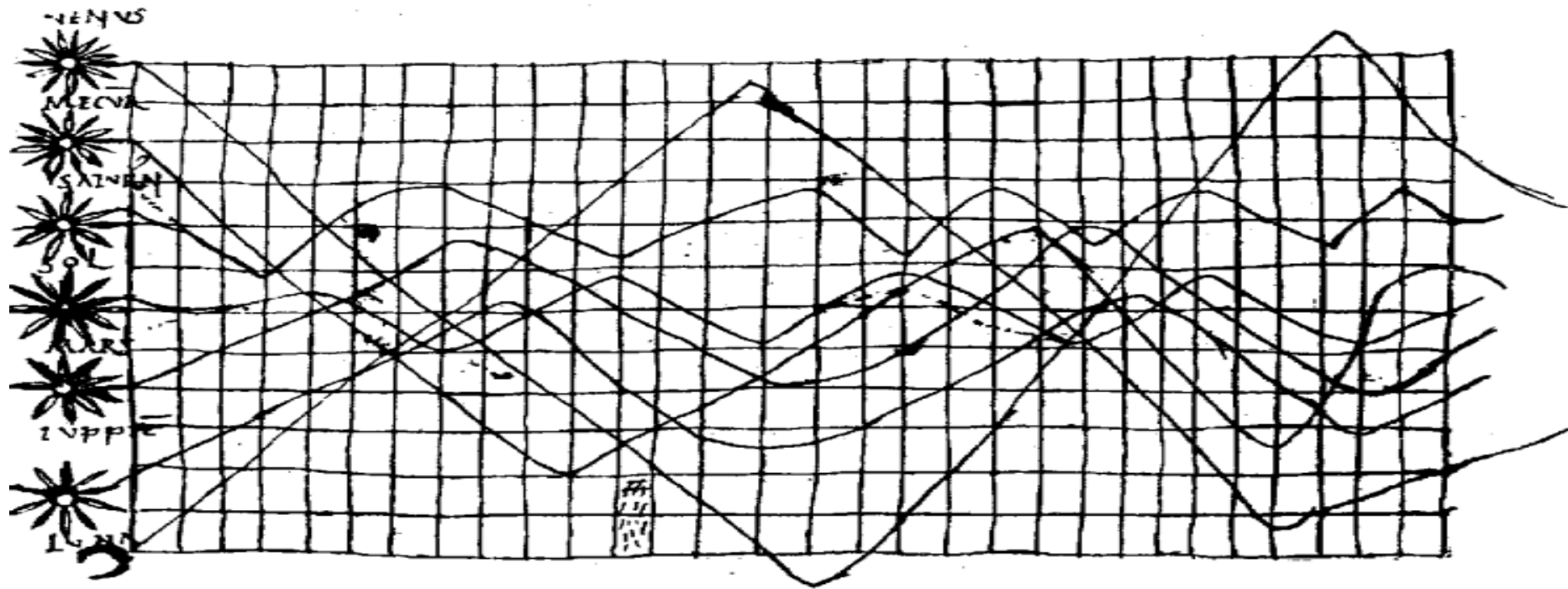
Data Visualization

Selected moments from history

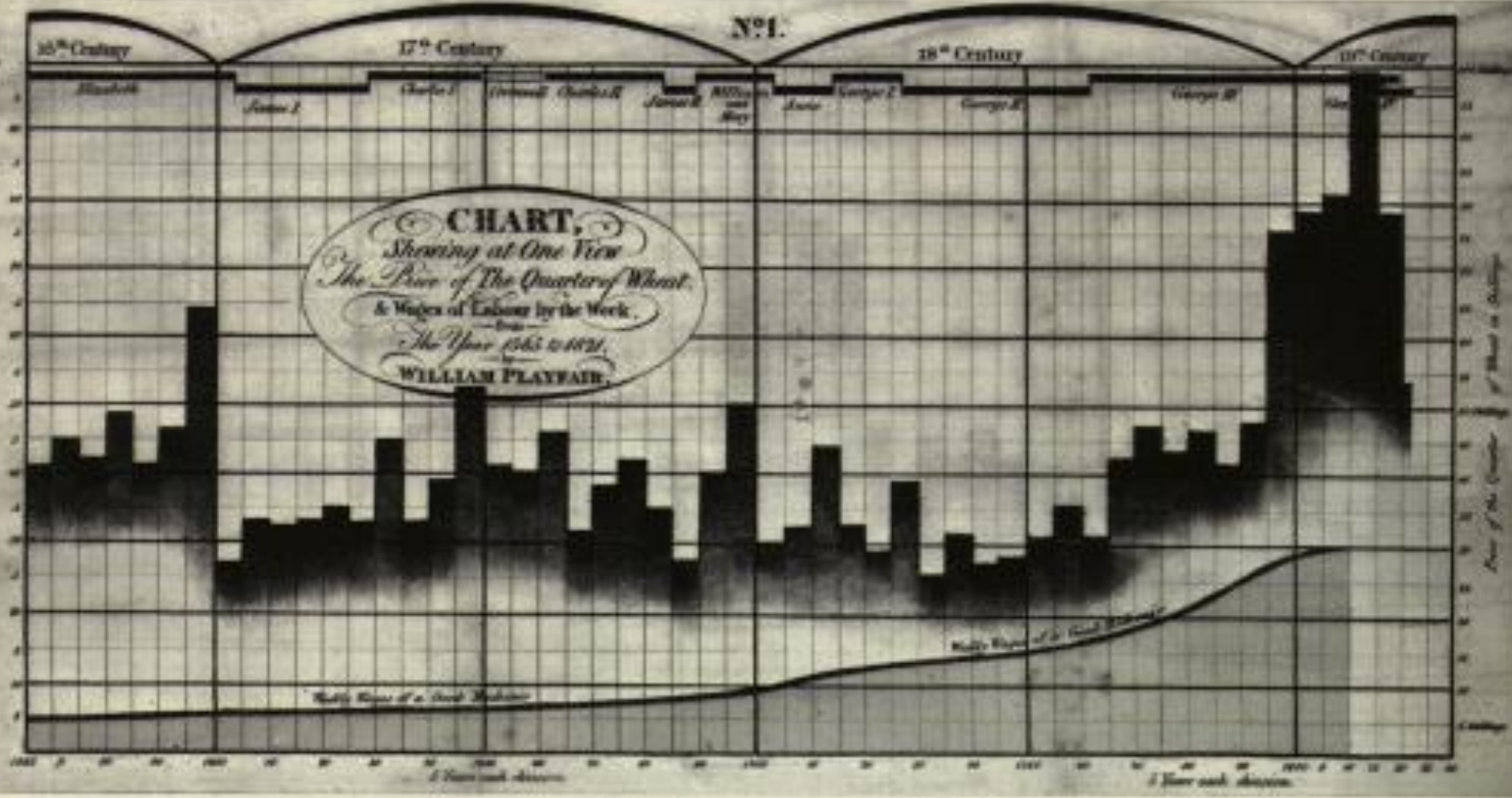




6 200 BC – Konya town map, Turkey



950 – planet movements

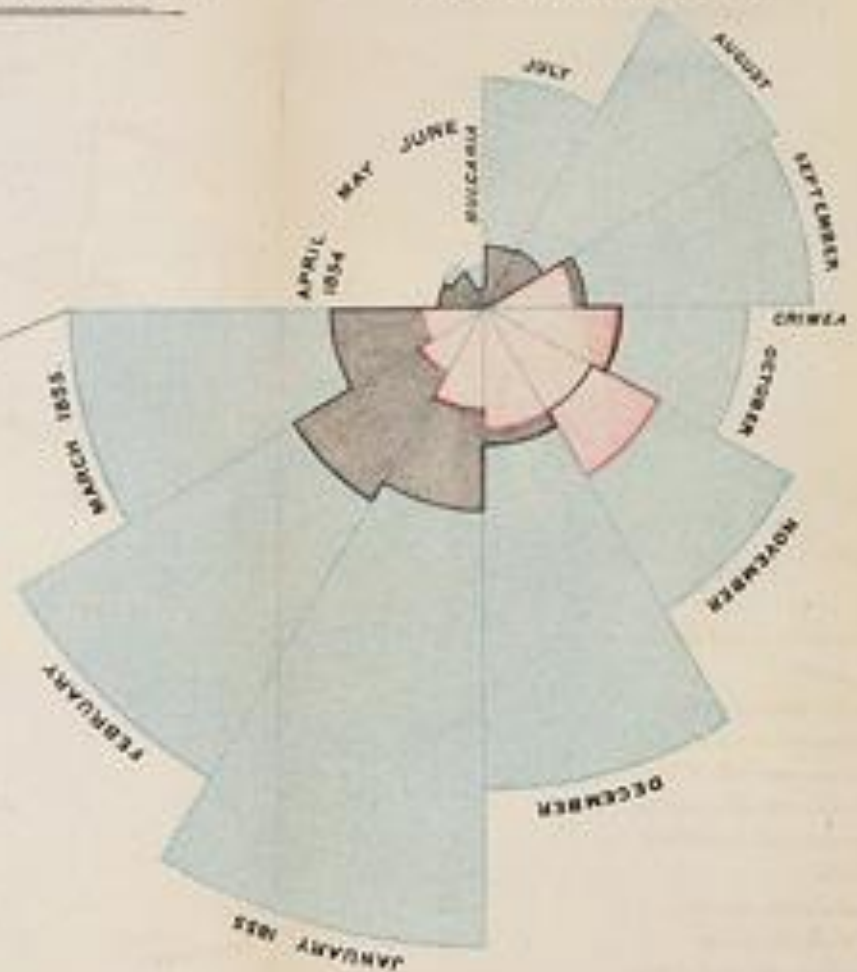
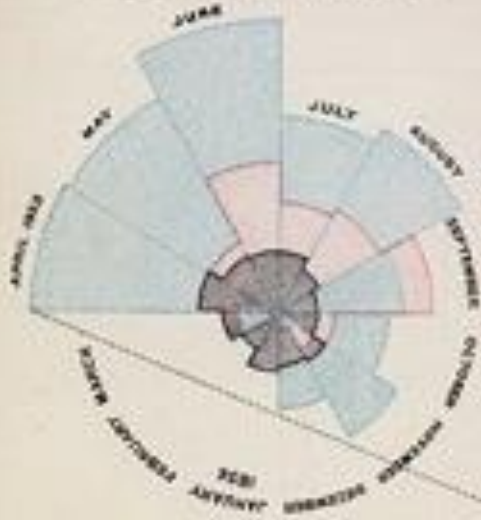


Bar chart, line graphs of economic data – price of wheat and wages
 William Playfair (1759-1823), England

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

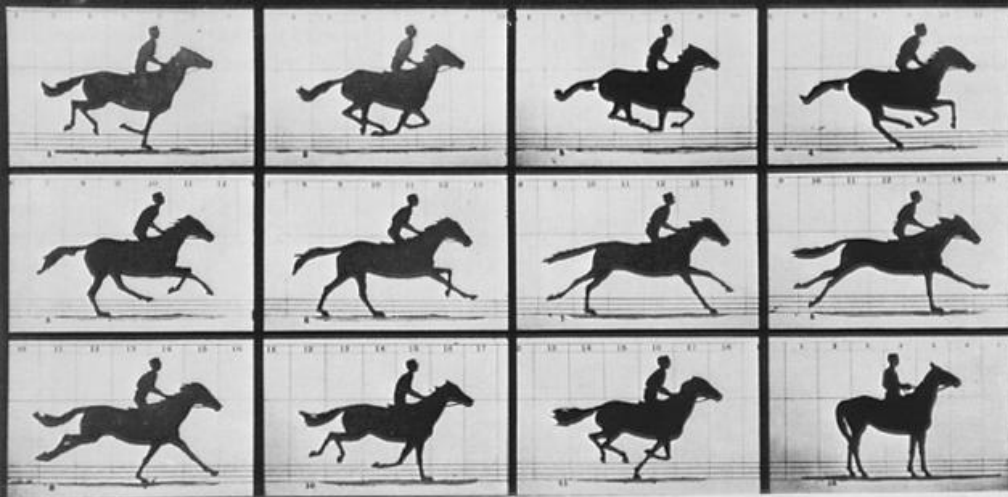
1
APRIL 1854 to MARCH 1855.

2
APRIL 1855 to MARCH 1856.



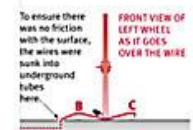
The areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
 The blue wedges measured from the centre of the circle represent area for area the deaths from Presentable or Malignant Zymotic diseases, the red wedges measured from the centre the deaths from wounds & the black wedges measured from the centre the deaths from all other causes.
 The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.
 In October 1854 & April 1855 the black area coincides with the red, in January & February 1855 the blue coincides with the black.
 The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Polar area charts, known as “coxcombs” used in a campaign to improve sanitary conditions of army; Florence Nightingale (1820-1910), England



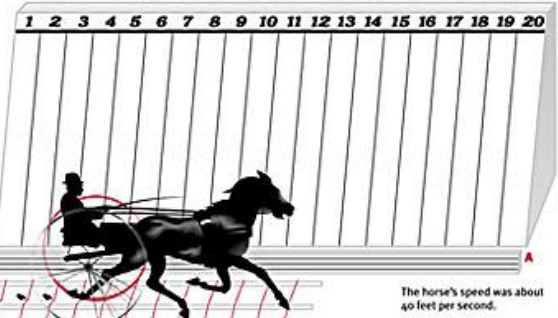
Muybridge's Photographic Method used on 15th June, 1878, at the Palo Alto Track.

Charles Martin, Leland Stanford's master trainer, steered the sulky pulled by champion trotter Abe Edgington so that the left wheel (shown in red) passed between two low strips of wood (B, C), across which were stretched fine wires.



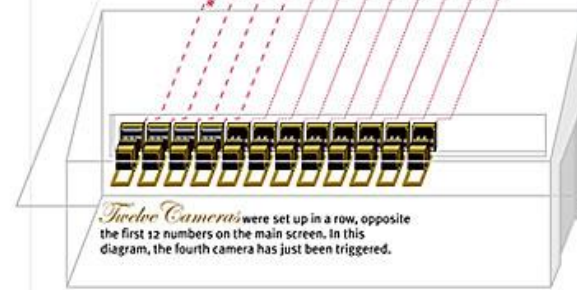
The raised wires were jerked down (⊖) by the sulky wheel passing over them. This triggered the completion of an electrical circuit, which in turn caused the shutter to work.

The Main Screen, set at a 20° angle and covered with white canvas, was marked with vertical black lines 21 inches apart. A much smaller screen (A) was set up in front, marked with horizontal lines four inches apart—to show the height of the horse's hooves above the ground.



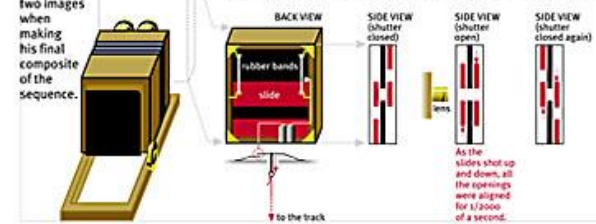
The Track was silted with powdered lime so that the surface appeared perfectly white in the photographs.

The Studio contained not only the cameras but also darkrooms, so that the photographs could be developed within minutes.



The Cameras each had two lenses. Muybridge selected the best of the two images when making his final composite of the sequence.

The Shutter Mechanism was housed in separate boxes that stood in front of the cameras. Two sets of slides (shown in red) passed in opposite directions across the front of the lenses. The slides were pulled by rubber bands that were released when an electric device anchoring the slides was triggered by the wires on the track.



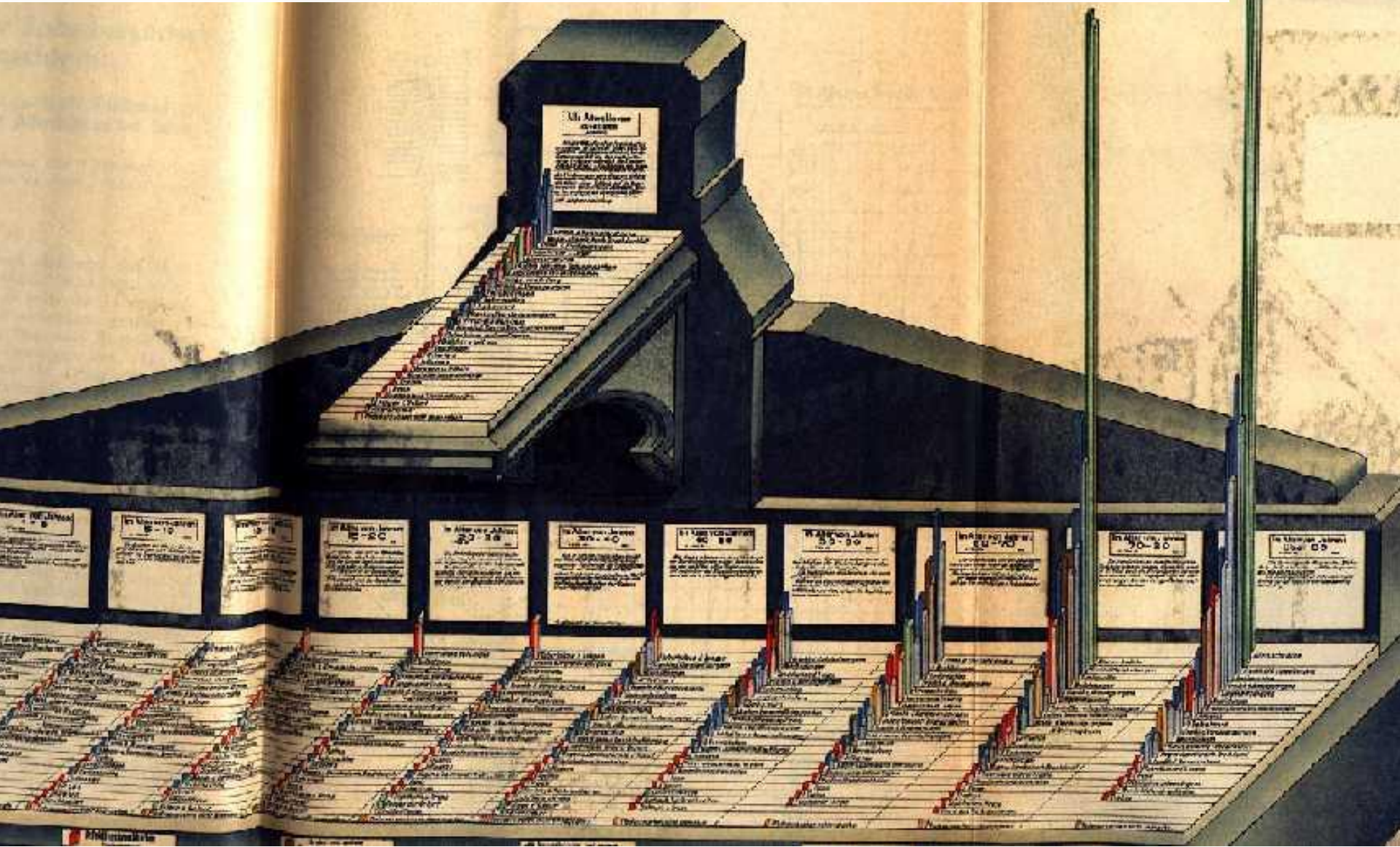
Later the same day, Muybridge photographed Stanford's racehorse Sallie Gardner. To do this, wires had to be stretched across the track at a height of about three feet, where the horse's chest would break them (⊖). This had the same triggering effect on the camera's shutter mechanism as the sulky wheel going over the wire on the ground. Within months, Muybridge developed a system of electrically timed shutter releases, which removed the need for any mechanical triggering. (It also allowed him to take pictures of birds in flight and other nonlinear movements.)

Copyright, U.S. by MUYBRIDGE. MORSE'S Gallery, 427 Montgomery St., San Francisco. **THE HORSE IN MOTION.** Illustrated by MUYBRIDGE. AUTOMATIC ELECTROPHOTOGRAPH. "SALLIE GARDNER," owned by LELAND STANFORD, running at a 140 gait over the Palo Alto track, 19th June, 1878. The negatives of these photographs were made at intervals of twenty-seven inches of distance, and about the twenty-fifth part of a second of time; they illustrate consecutive positions assumed in each twenty-seven inches of progress during a single stride of the mare. The vertical lines were twenty-seven inches apart; the horizontal lines represented elevations of four inches track. The exposure of each negative was less than one two-hundredth part of a second.



Recording of motion (of a running horse) by means of a set of glass-plate cameras, triggered by strings. Eadweard Muybridge (1830-1904), USA

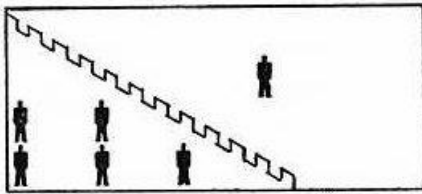
First International Hygiene-Exhibition in Dresden, 1911
with 259 graphical-statistical figures of 35 exhibitors and more than 5 million
visitors. Germany
3D Histogram: The course of death in Saxony



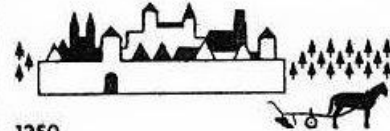
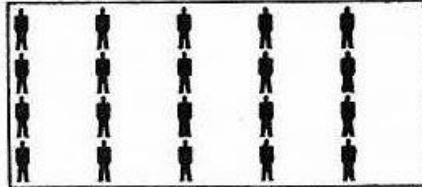


Parade of statistical graphics, May 17, 1913, including large graphs on horse-drawn floats, and a photograph with people arranged in a bell-shaped curve, New York, USA

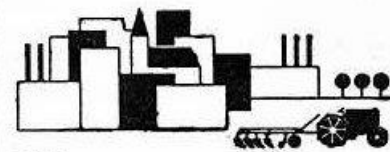
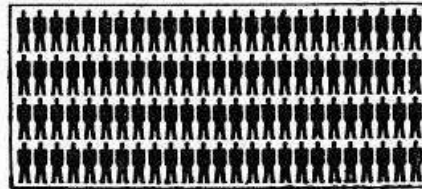
Number of Men Living in Europe



In the year of Christ's birth



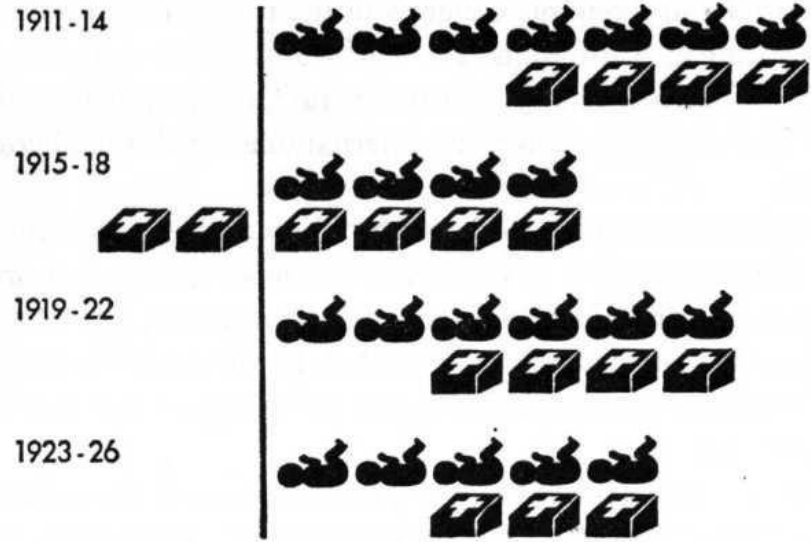
1350



1930

1 sign for 5,000,000 men

Births and Deaths in Germany in a Year

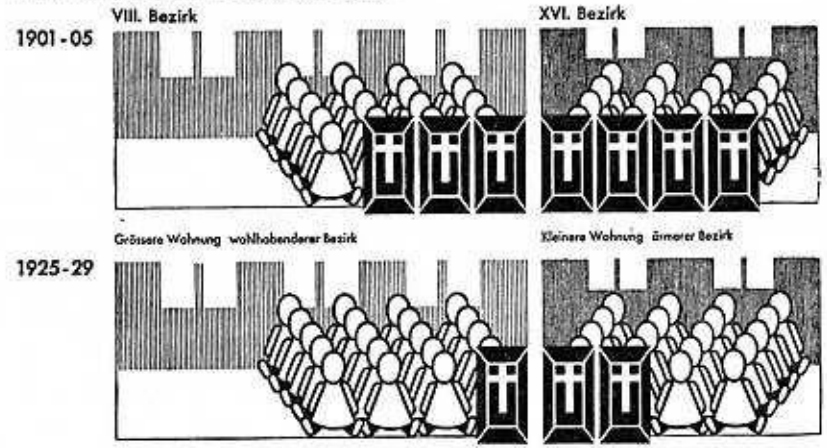


1 child for 250,000 births a year
1 coffin for 250,000 deaths a year

Otto Neurath (Director) (1882-1945),
Social and Economic Museum, Vienna,
Austria

Museum of Social Statistical Graphics and
the ISOTYPE system

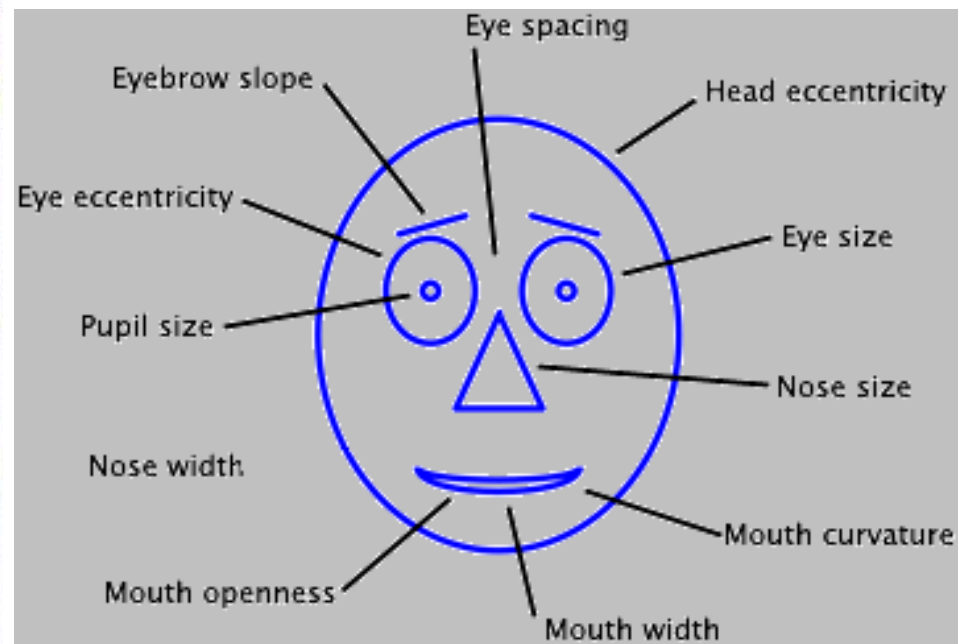
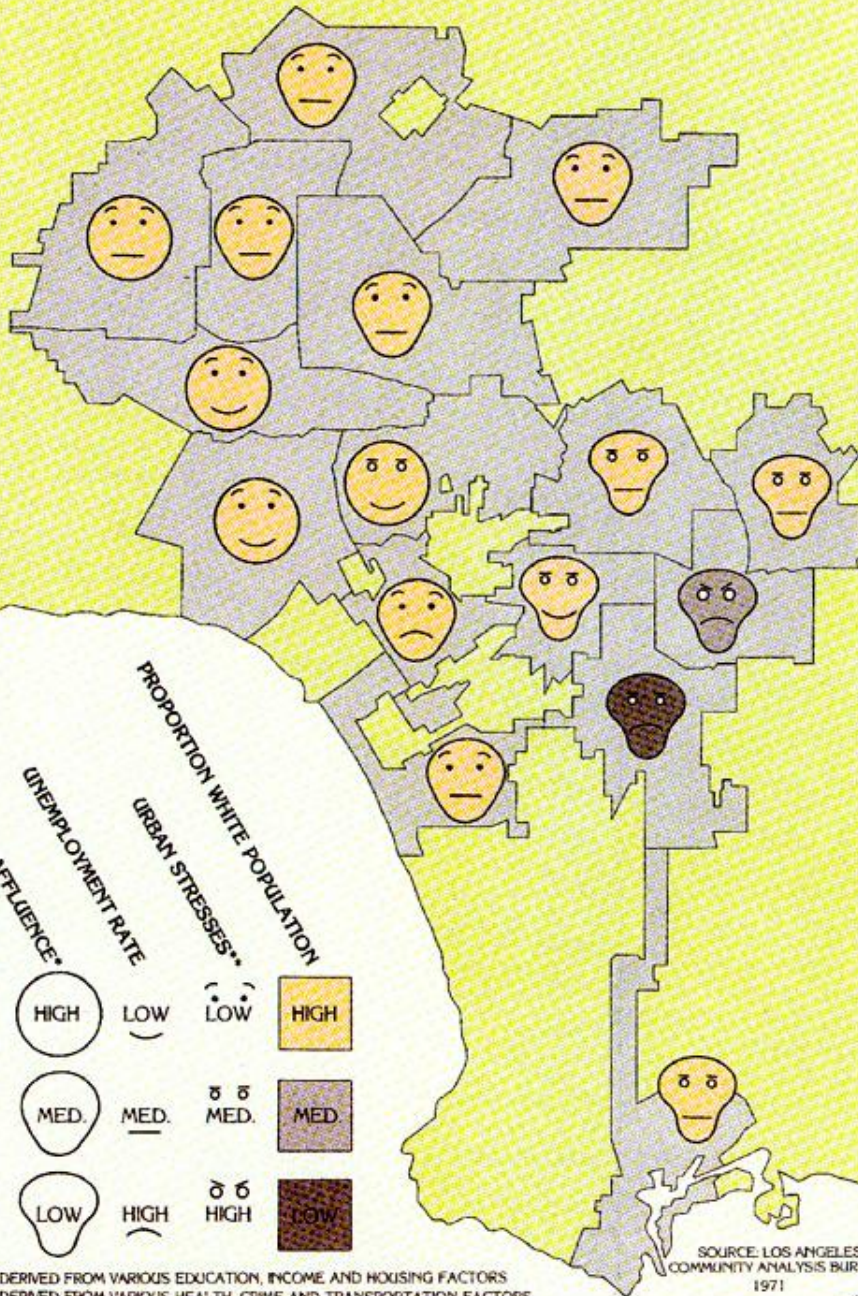
Säuglingsterblichkeit und soziale Lage in Wien
Infant Mortality and Social Position in Vienna



Todesfälle im ersten Lebensjahr auf 20 Lebendgeburten
Deaths in the first year of life out of every 20 children born alive

Gesellschafts- und Wirtschaftsmuseum
in Wien

Life in Los Angeles



In 1973, Herman Chernoff introduced a visualization technique to illustrate trends in multidimensional data

<http://kspark.kaist.ac.kr/HumanEngineering.files/Chernoff/ChernoffFaces.htm>



John Wilder Tukey

LIFE

Born June 16, 1915
New Bedford, Massachusetts, USA

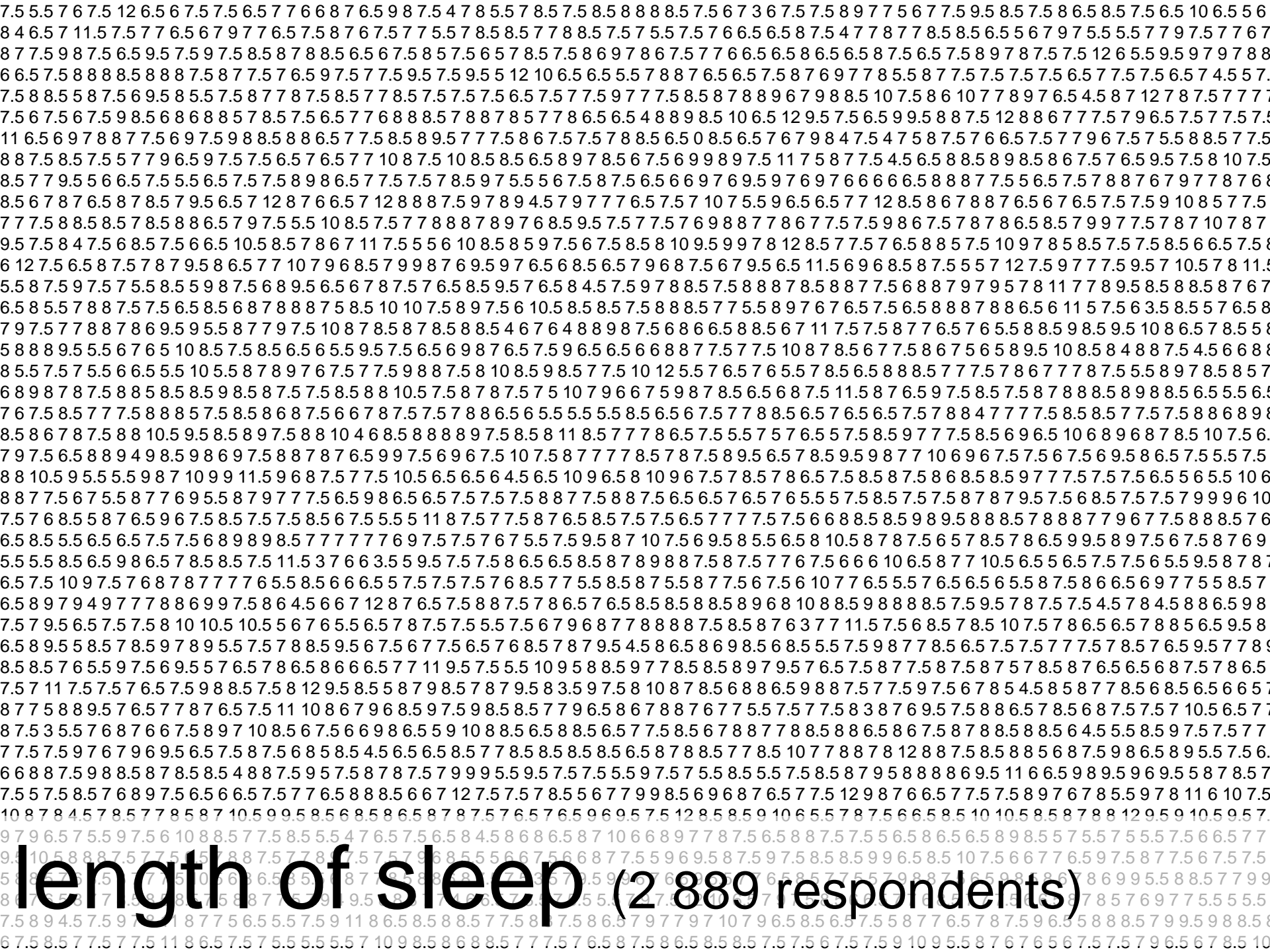
Died July 26, 2000 (aged 85)
New Brunswick, New Jersey

Known for FFT algorithm
Box plot
Coining the term 'bit' (binary digit)
Exploratory data analysis (EDA)

Tukey, John Wilder (1962).
The future of data analysis.
Annals of Mathematical Statistics, 33:1-67 and 81.

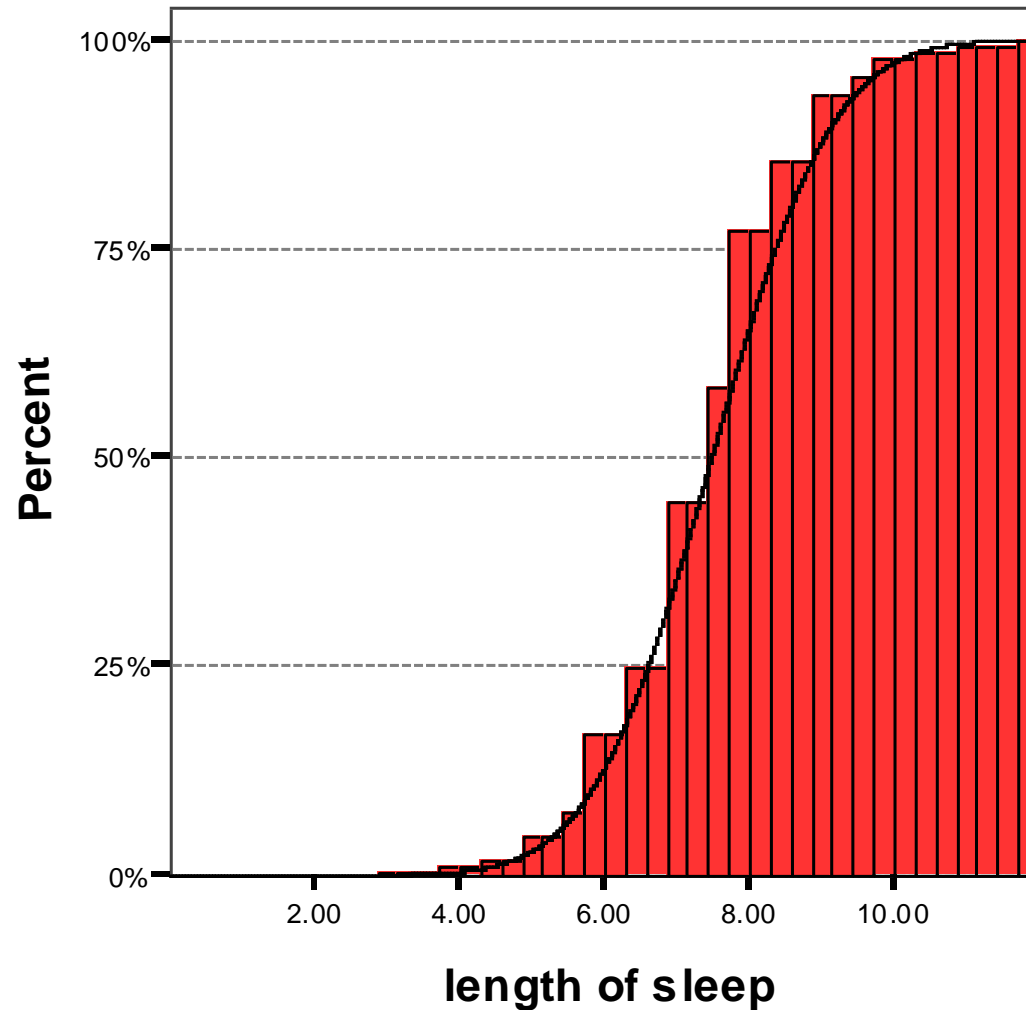
Univariate charts



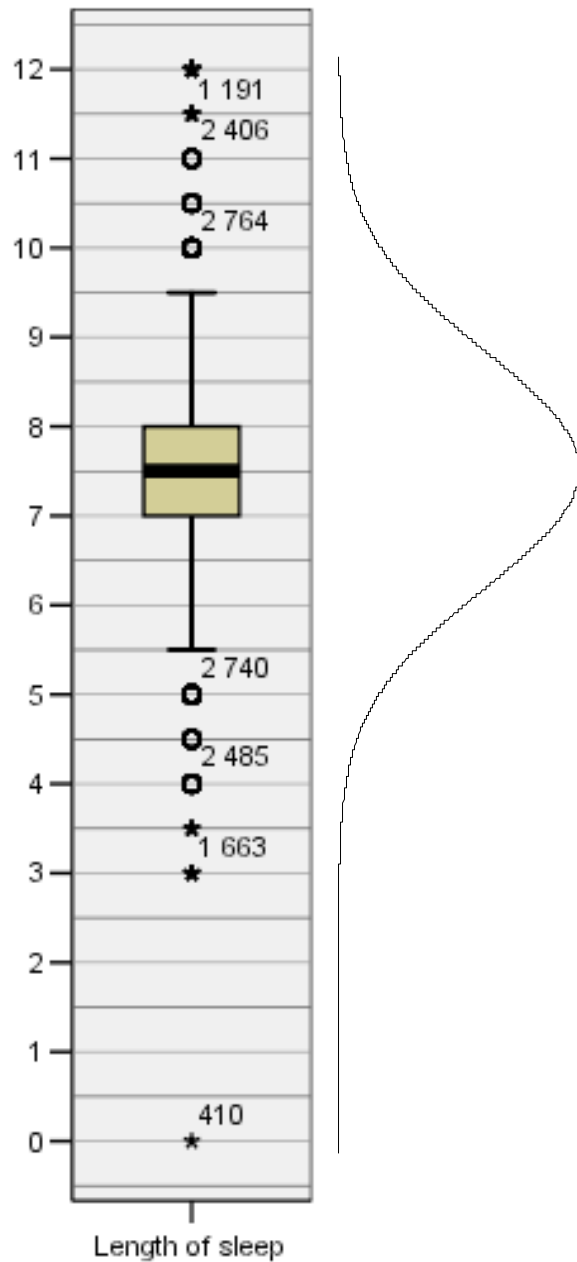


Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
denspa Length of sleep	2889	.00	12.00	7.4673	1.30312
Valid N (listwise)	2889				

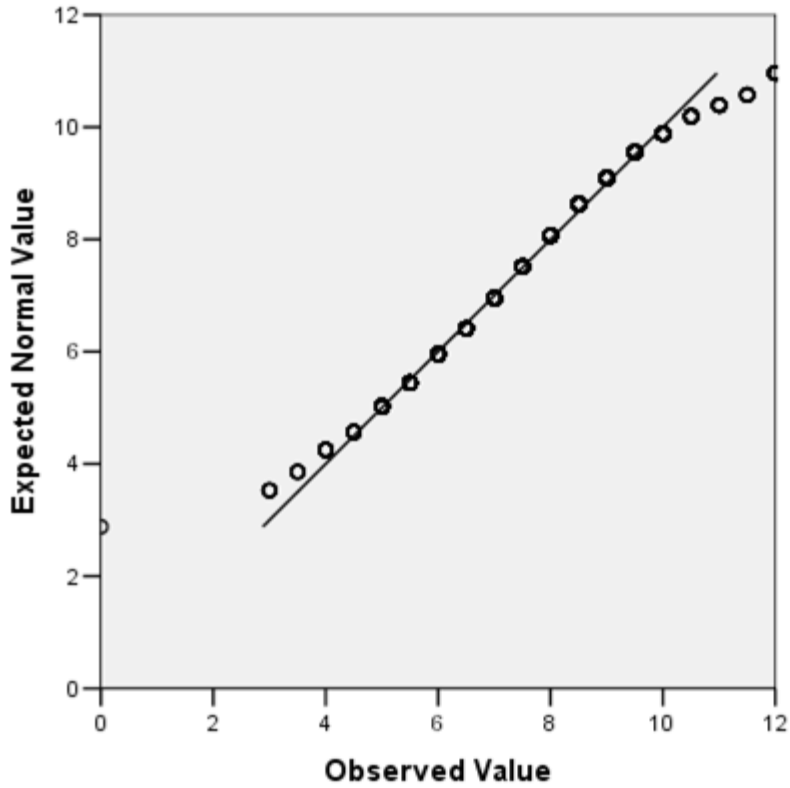


Cumulative histogram
(distribution estimator) of
length of sleep among
respondents

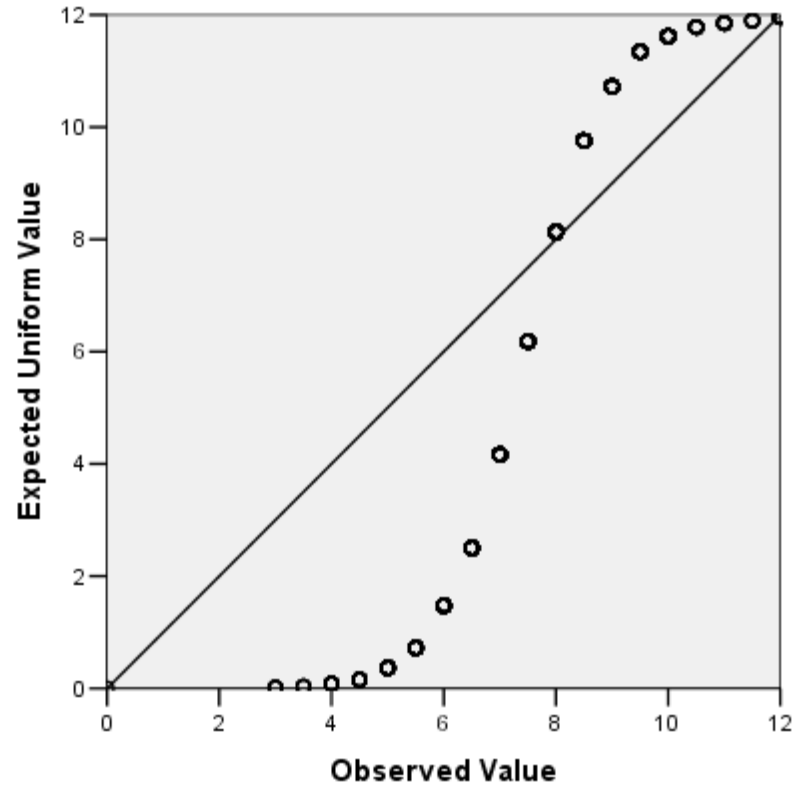


Box plot of length of sleep among respondents

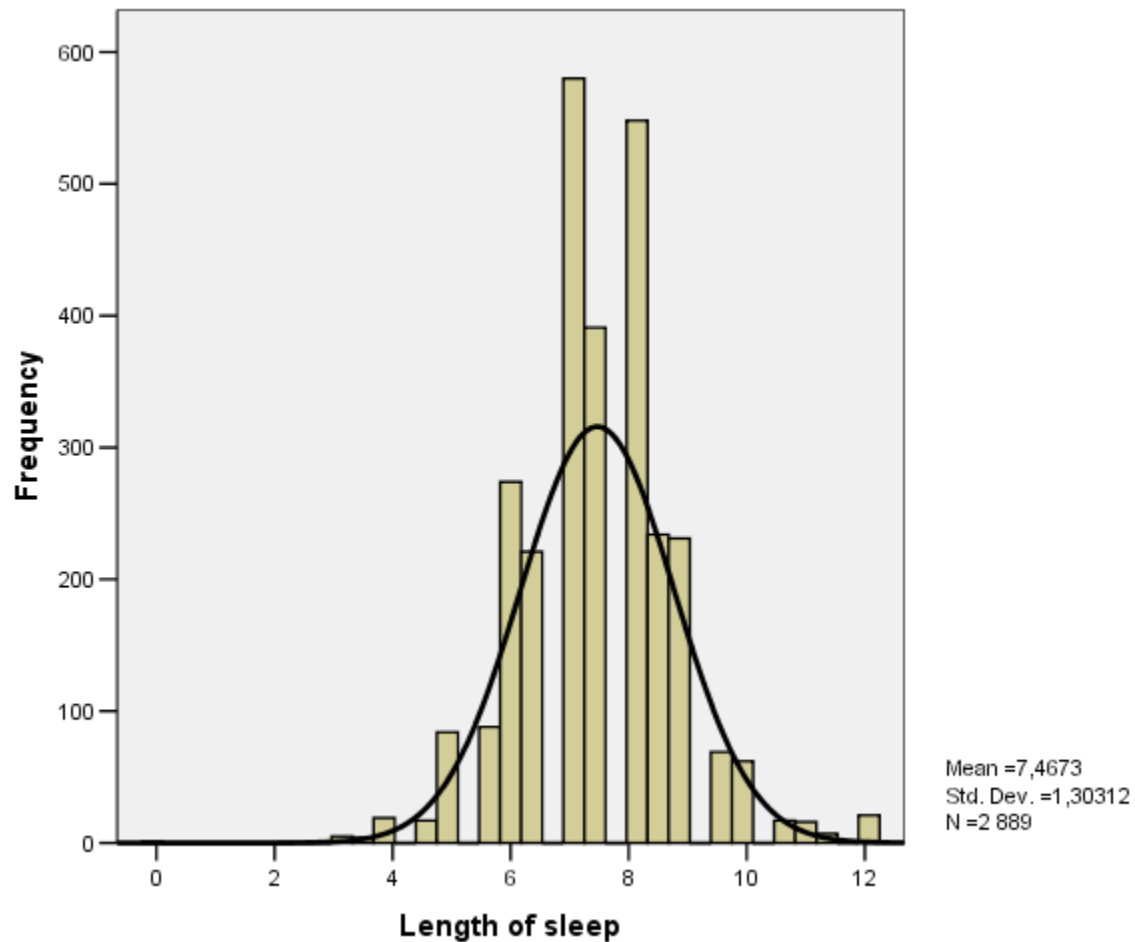
Normal Q-Q Plot of Length of sleep



Uniform Q-Q Plot of Length of sleep



Normal q-q plot of length of sleep among respondents

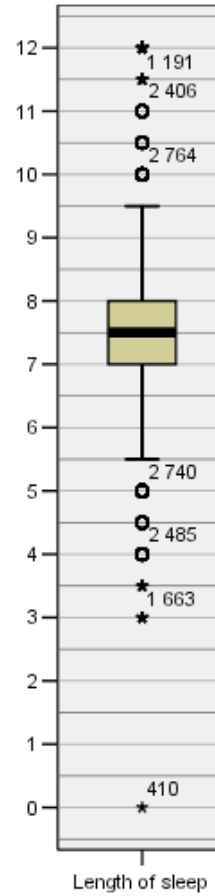
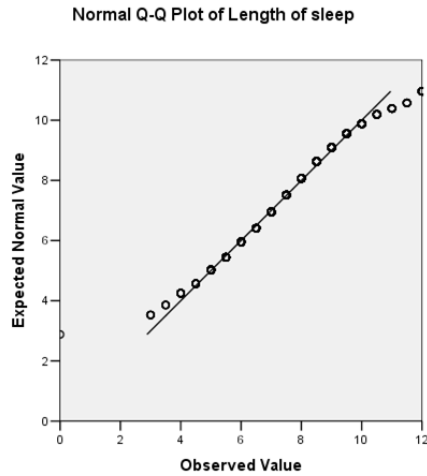
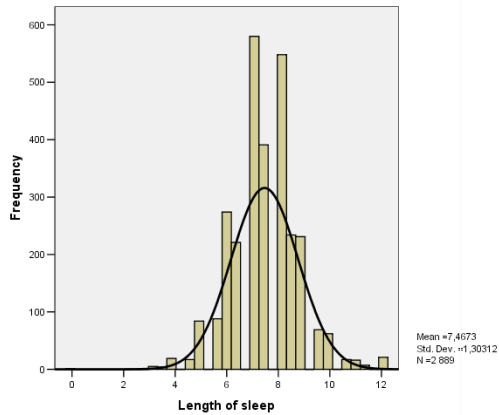


Histogram of length of sleep among respondents



we sleep about

7.5 hours a day



Descriptive Statistics

	Minimum	Maximum	Mean	Std. Deviation
denspa Length of sleep	.00	12.00	7.4673	1.30312
Valid N (listwise)	2889	2889		

$$\emptyset = 7.5$$

complexity

demanding

information

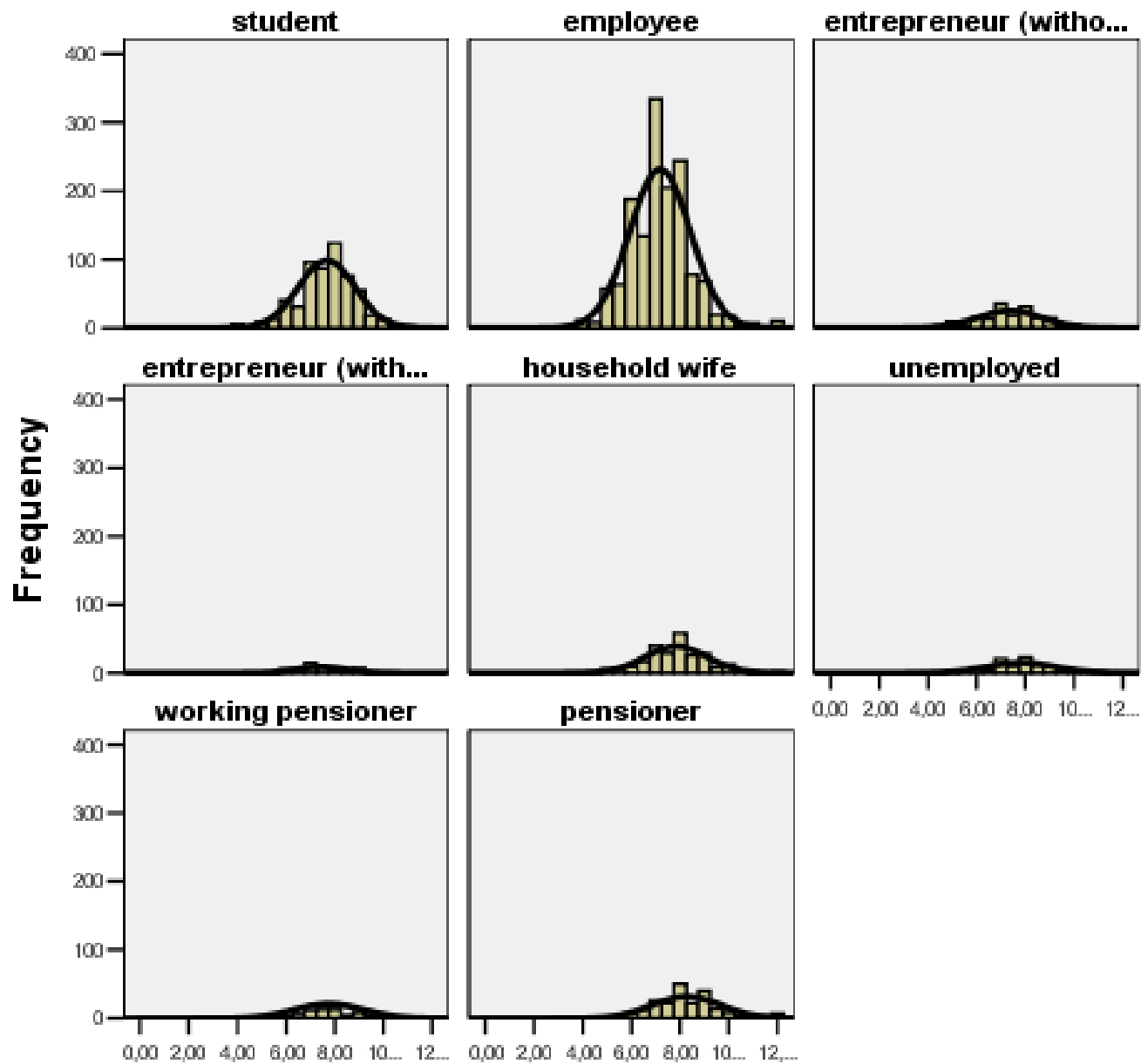
simplicity

self-explaining

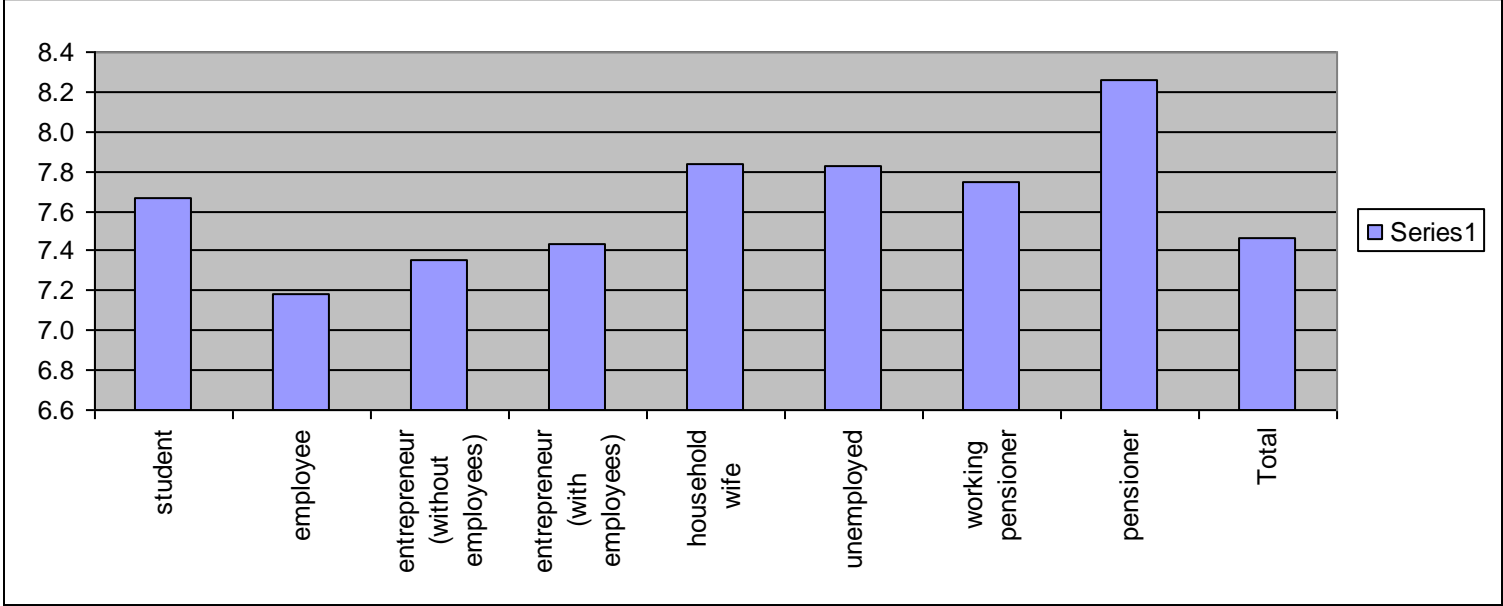
knowledge

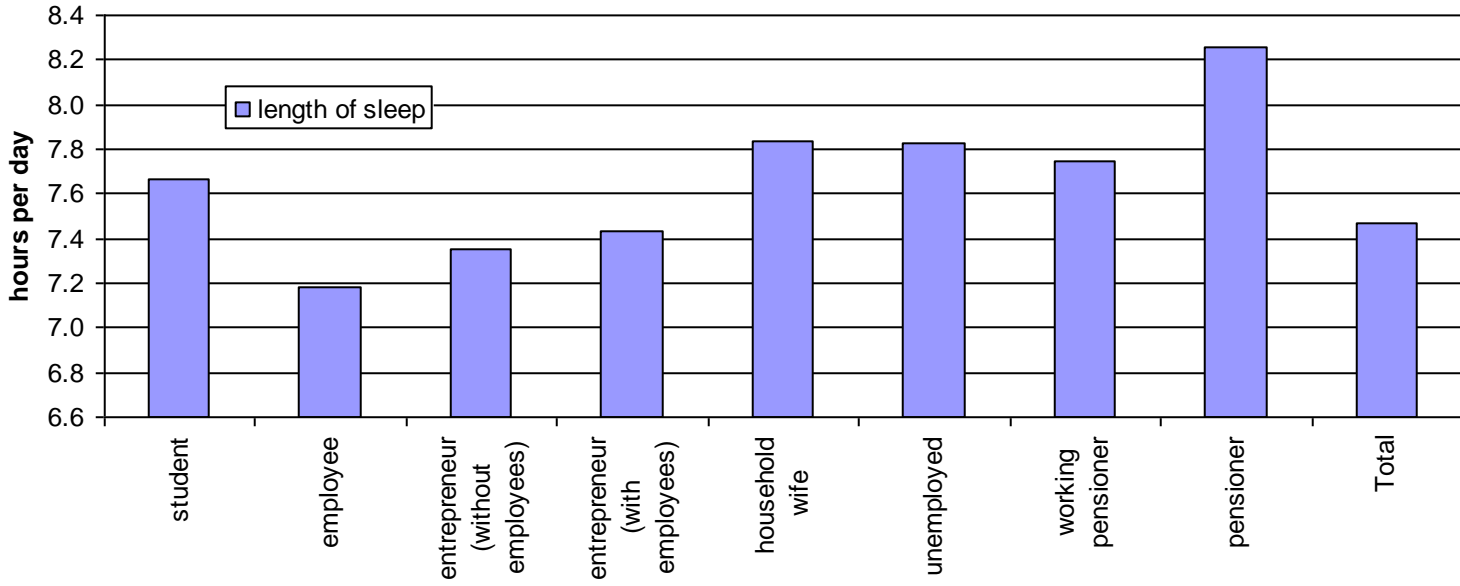
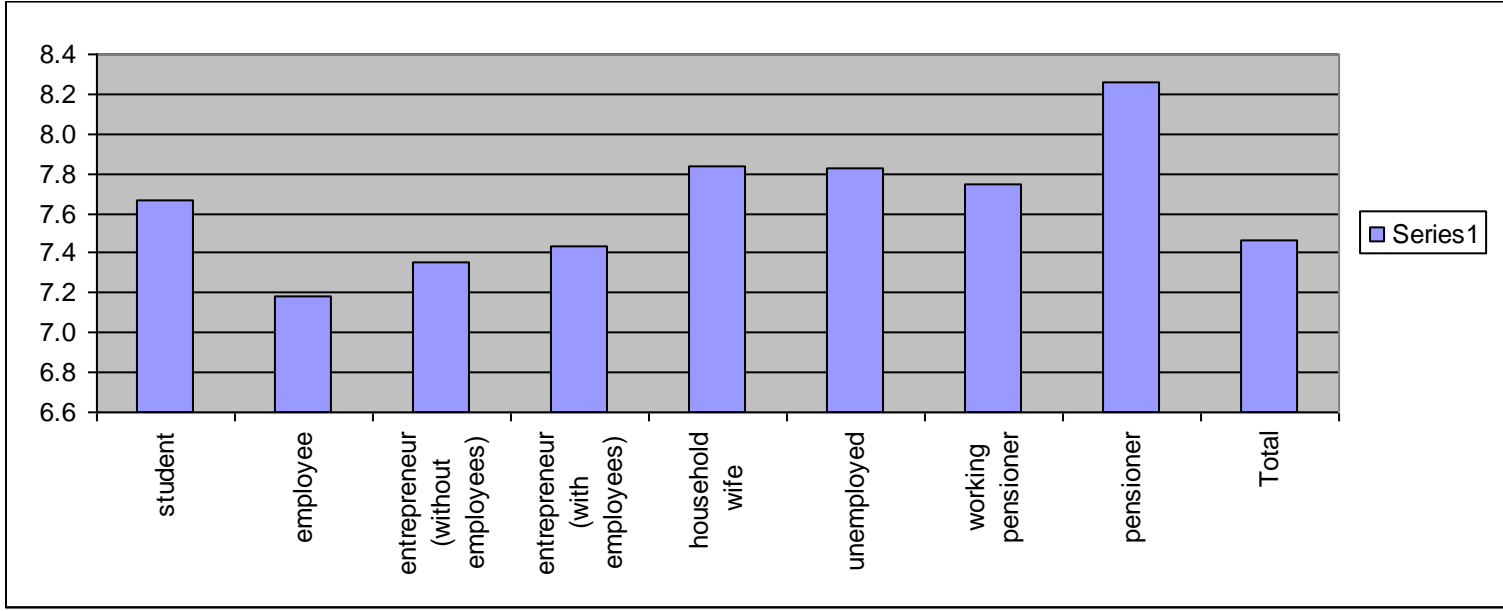


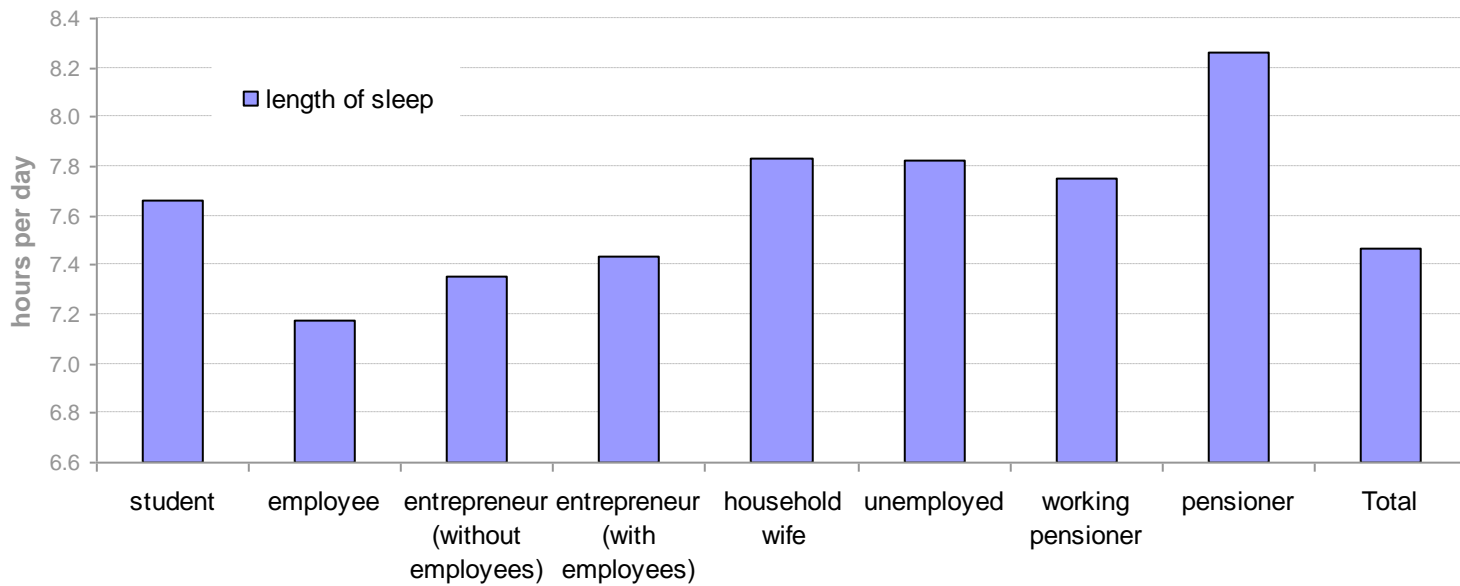
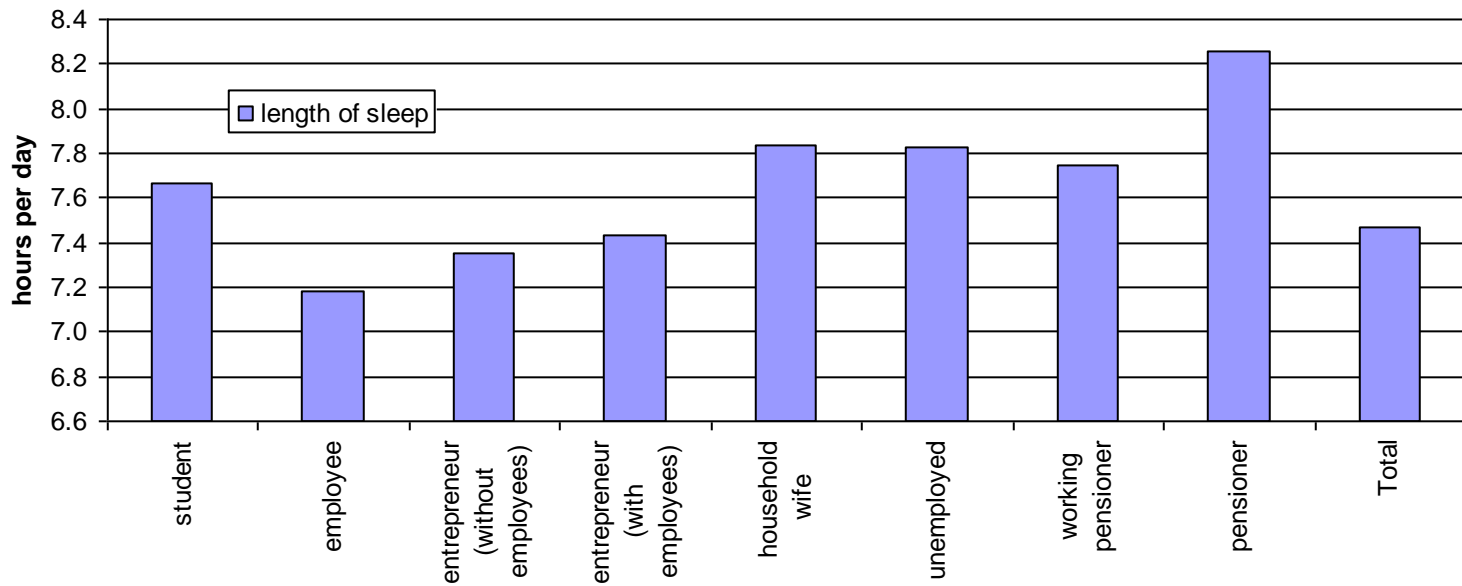
Bivariate charts

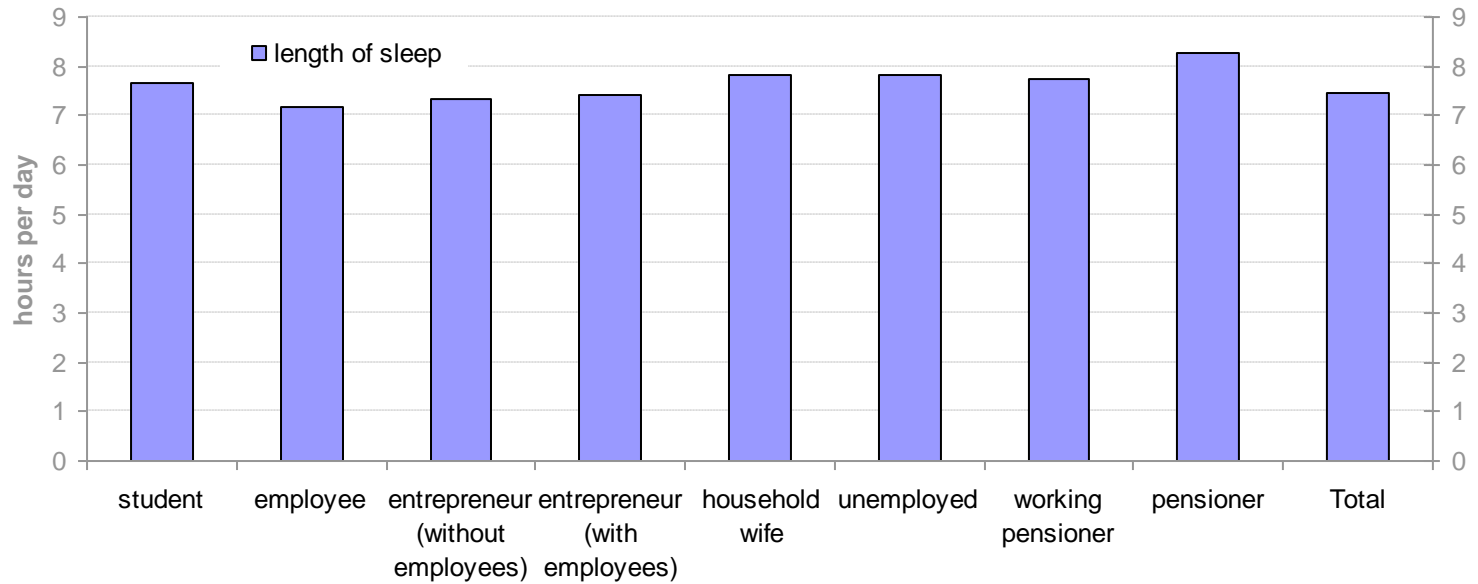
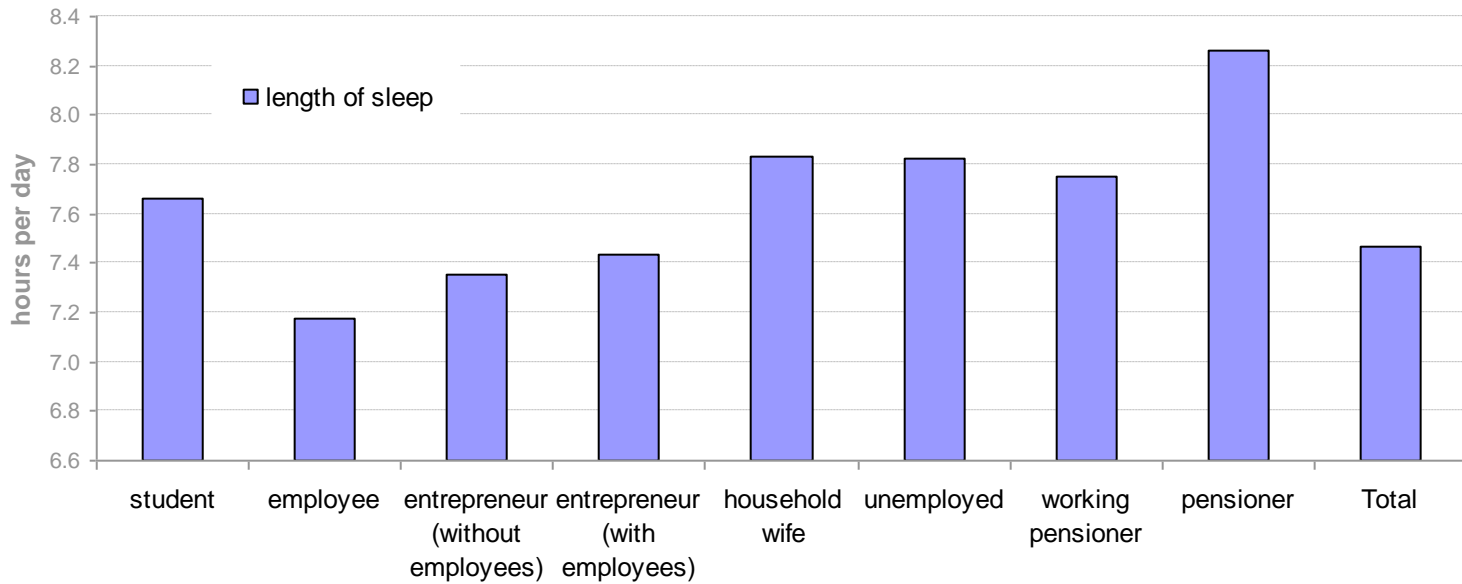


Histogram of length of sleep across job status (trellis plot, coplot)

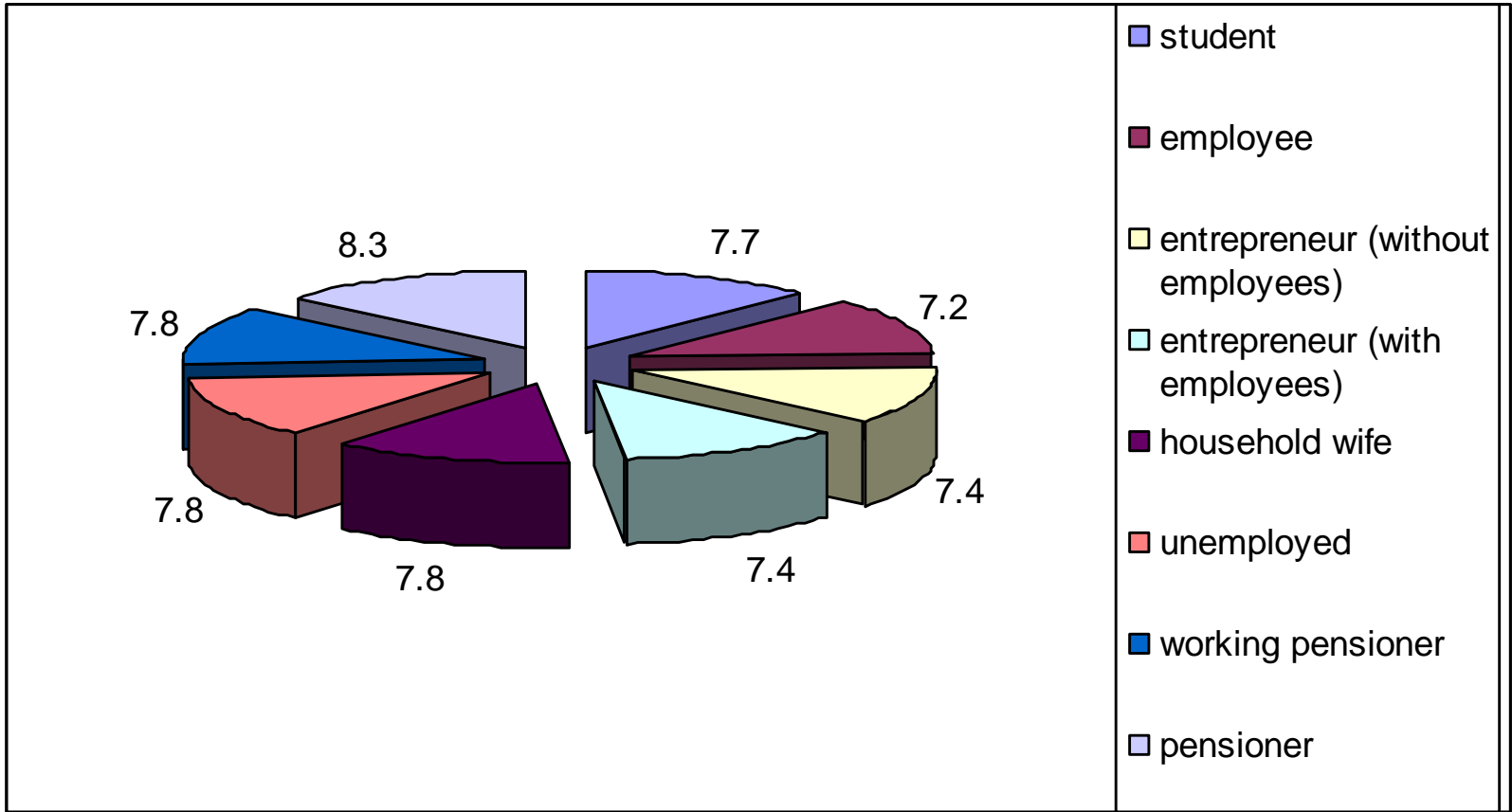






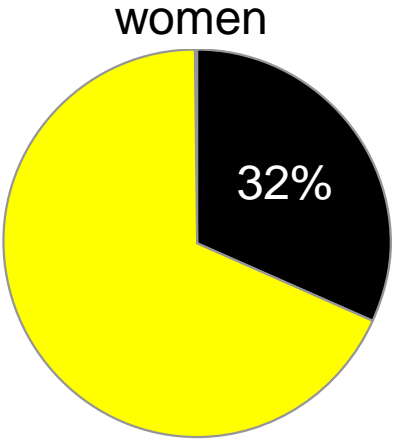
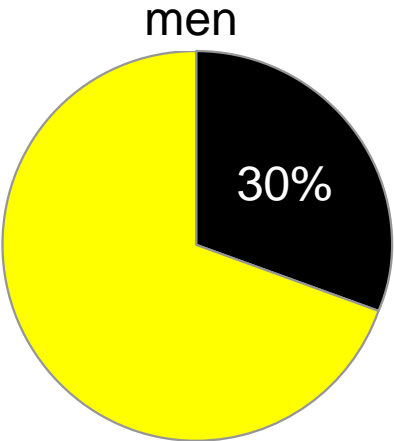


How long do *y*OU sleep?

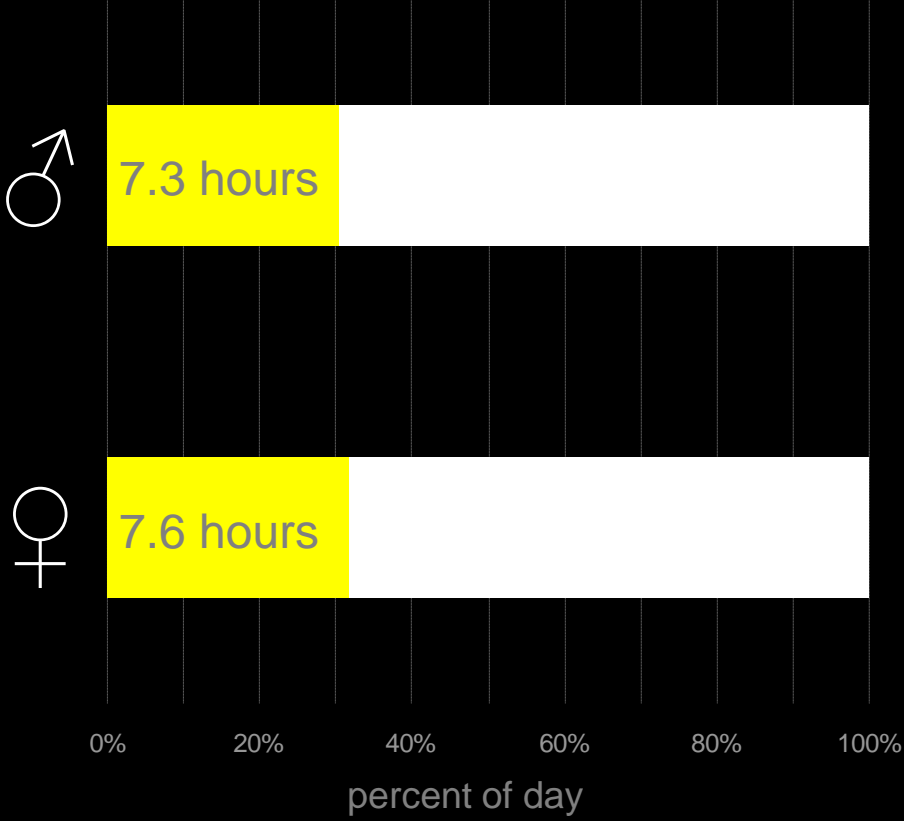


3D pie-chart of length of sleep across job status

Average length of sleep (% of day):

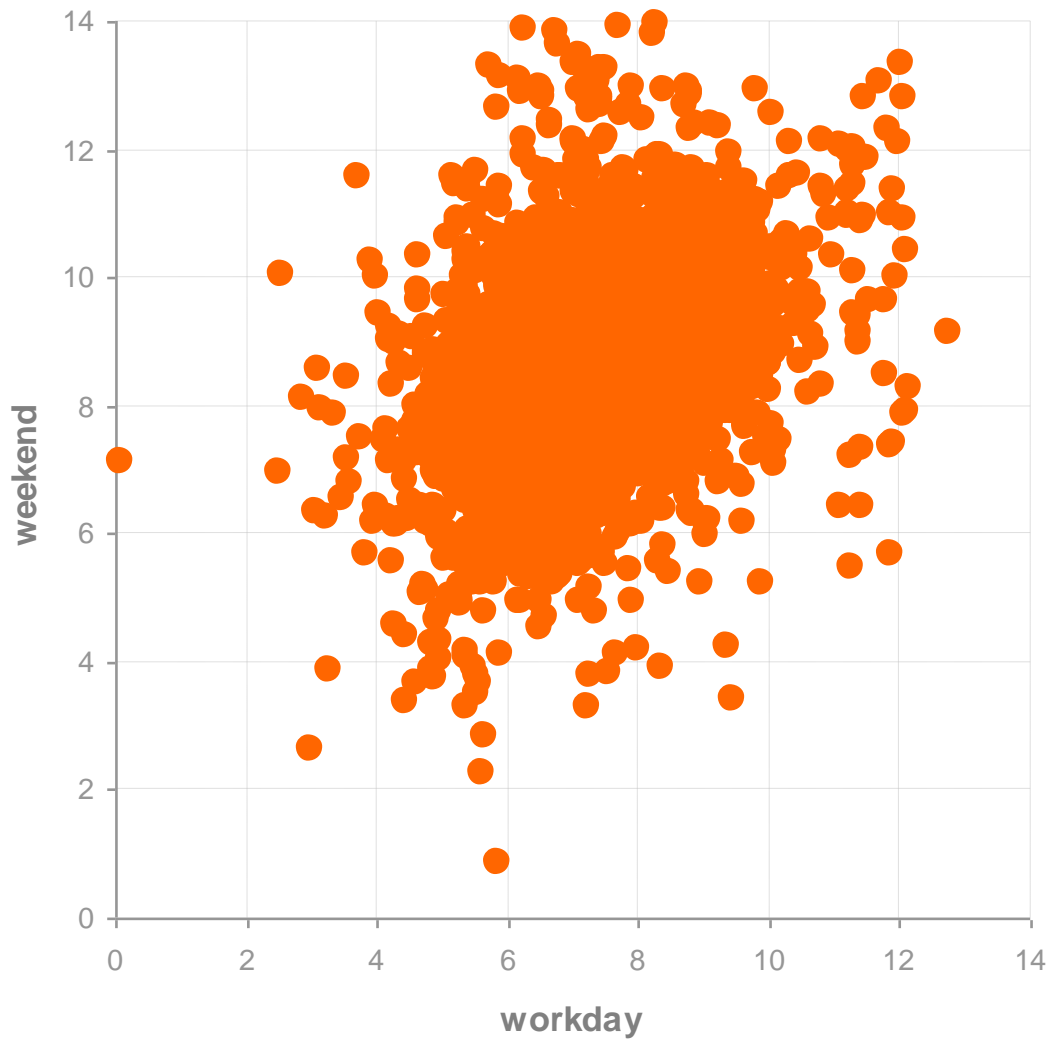


Average length of sleep

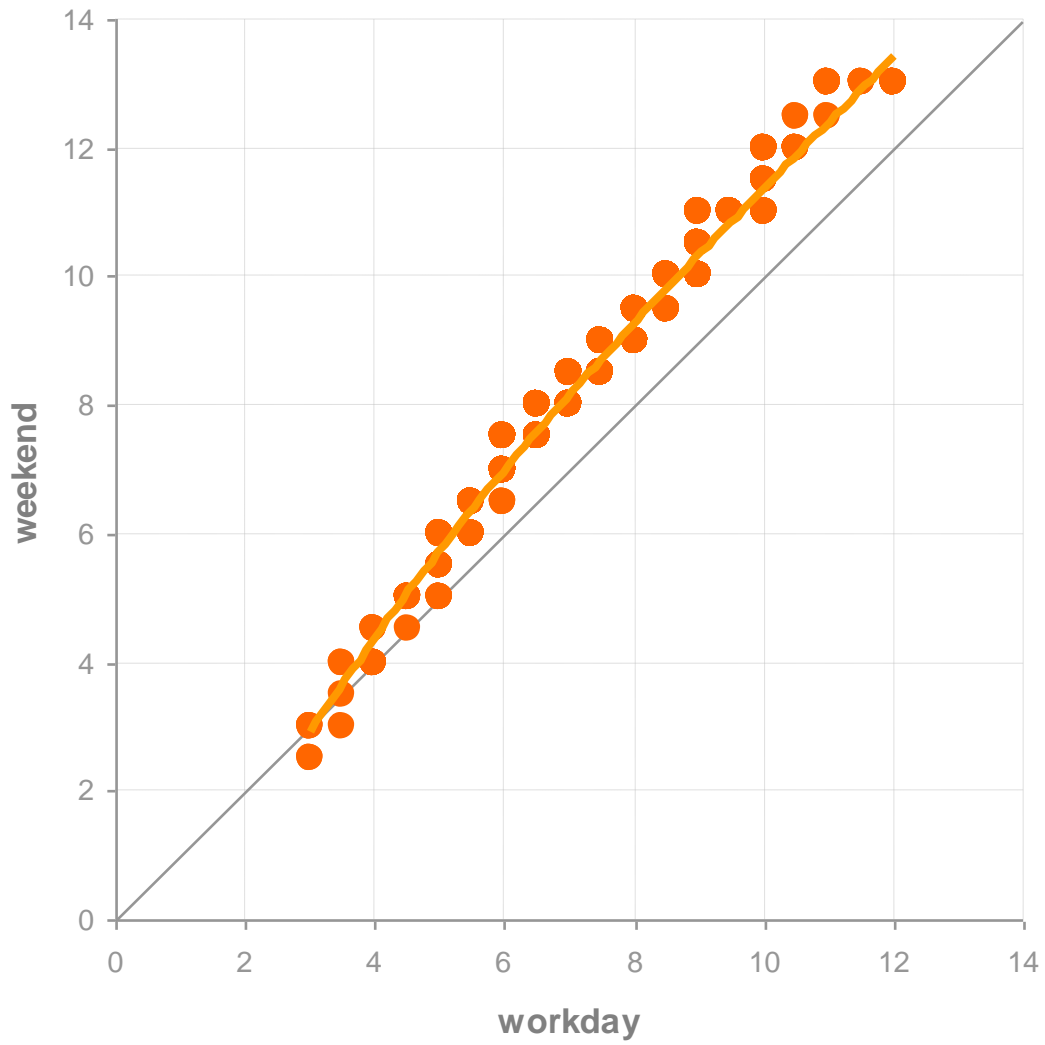




Scatter plot of length of sleep – workday vs. weekend



Jittering of points in scatter plot



Q-Q plot

Tableau figuratif du mouvement commercial du Canal du Centre en 1844

dressé par M^r. Minard sur les renseignements de M^r. Comoy. le 6^{ra} 1845 Ch. J. Minard

Le mouvement total équivalent à 131,000 tonneaux parcourant la longueur du Canal ou 117 kilomètres.

Le transit y est compris pour 10,000 tonneaux.

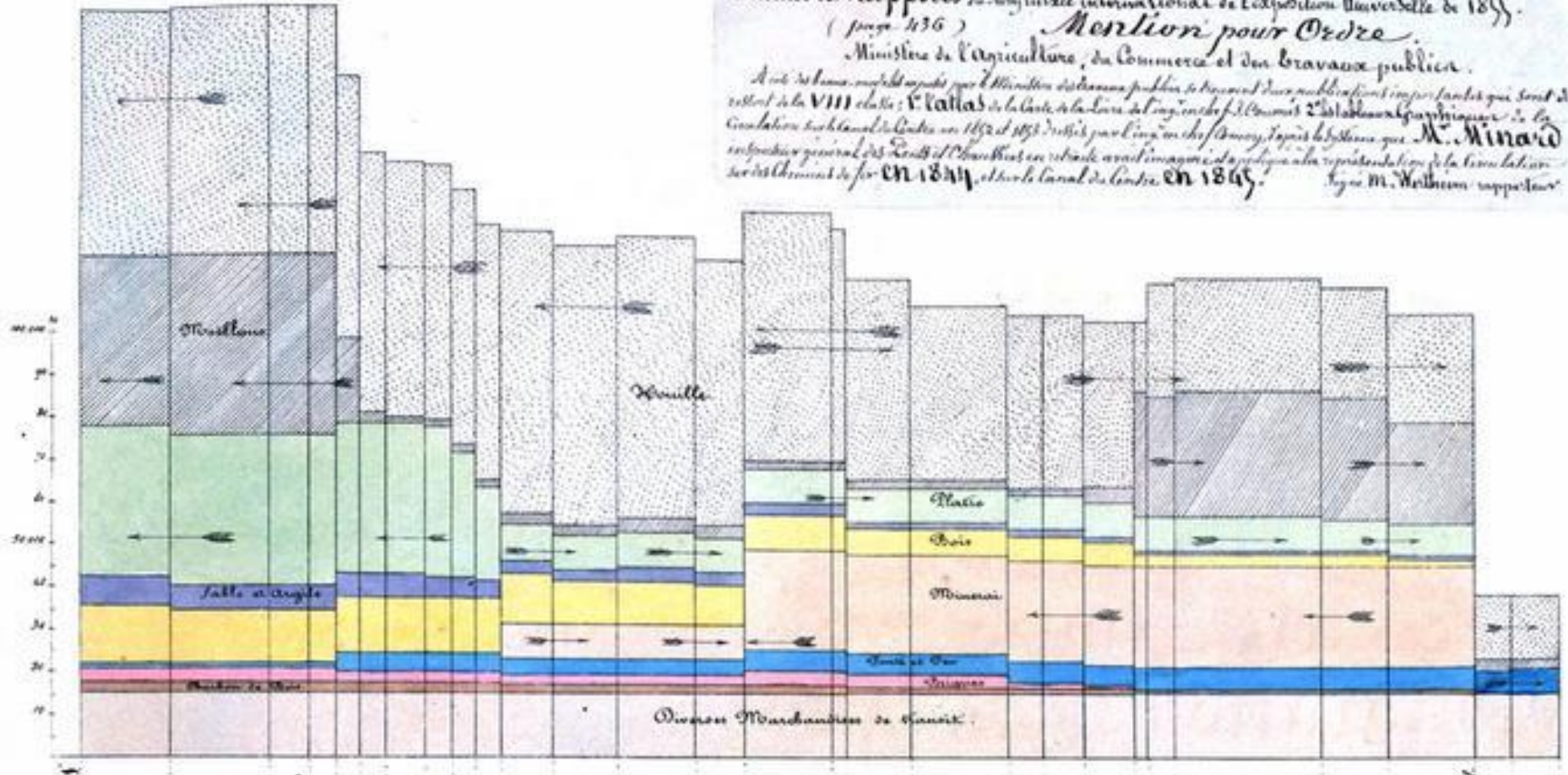
Extrait des Rapports du Congrès international de l'exposition universelle de 1855.

(page 136)

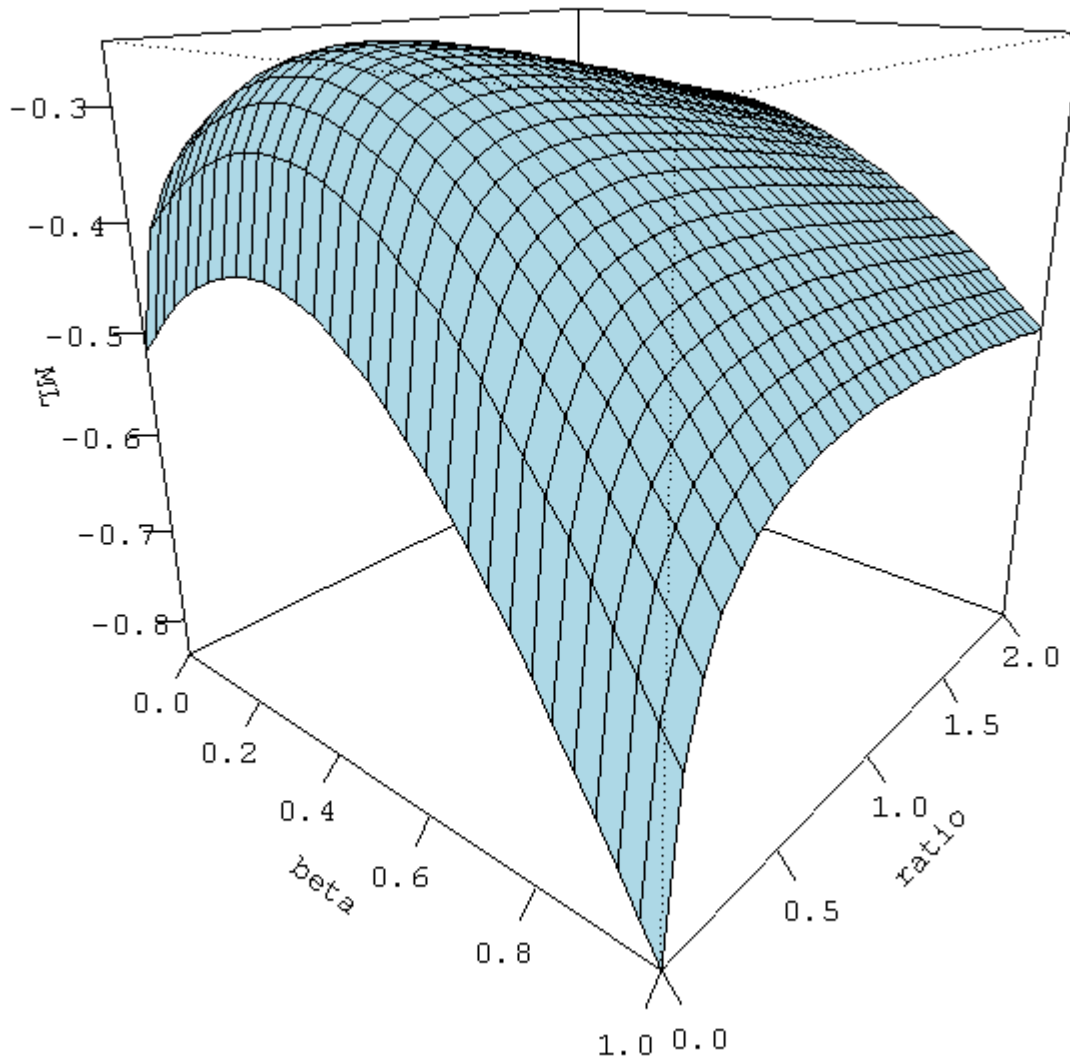
Mention pour Ordre.

Ministère de l'Agriculture, du Commerce et des Travaux publics.

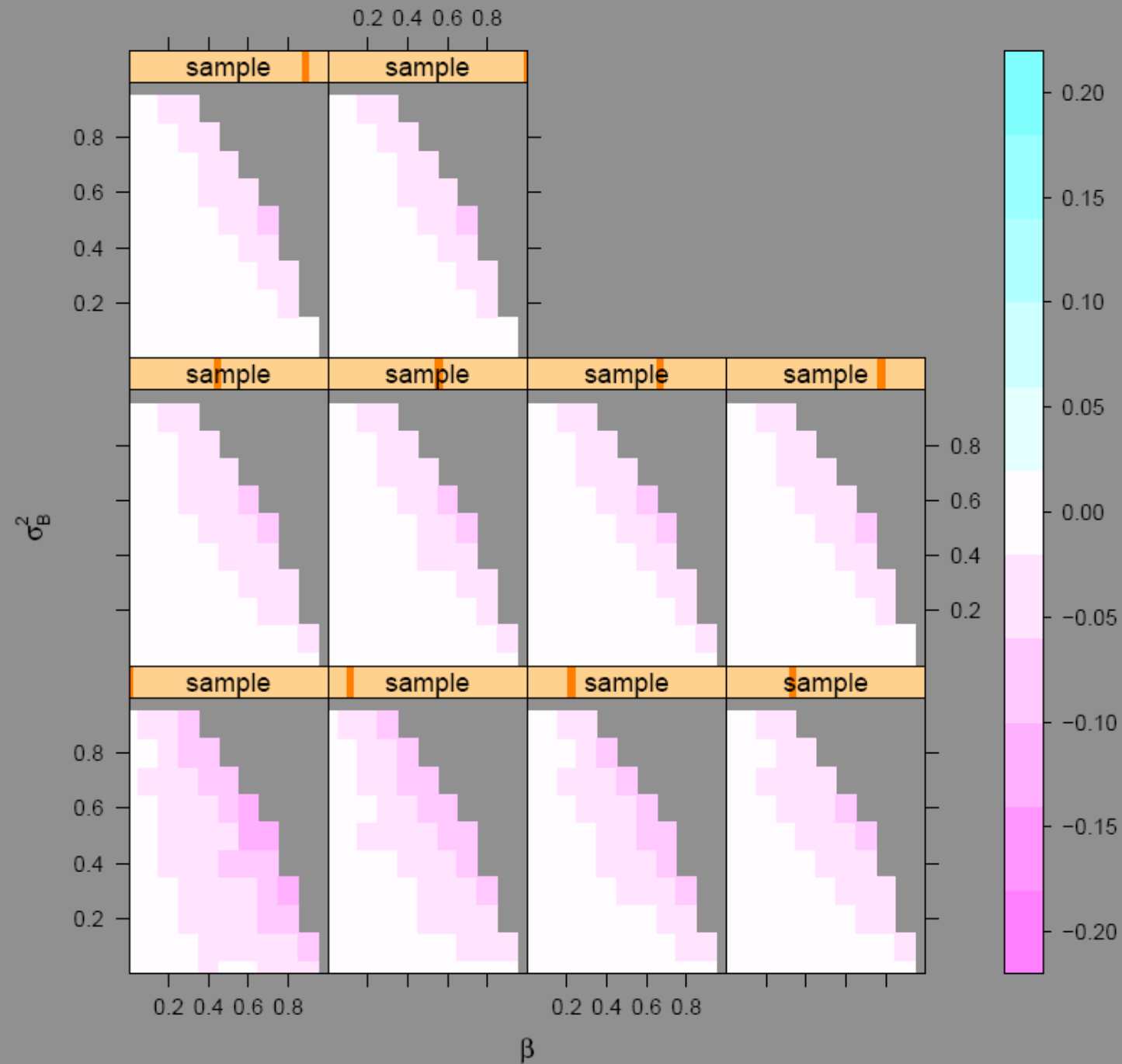
A cet égard, nous devons mentionner que le Ministre des Travaux publics a fait publier deux publications importantes qui sont devenues de la VIII^e classe: l'Atlas de la Carte de la Loire et l'Atlas de l'Établissement par le Canal de la Loire en 1852 et 1853 dressés par l'ingénieur en chef Comoy, depuis lesquels M. Minard a pu faire un grand nombre de travaux en ce genre, et il a été heureux de voir son nom mentionné dans les Rapports et dans les programmes de l'exposition universelle de 1855. Signé M. Wertheim rapporteur.



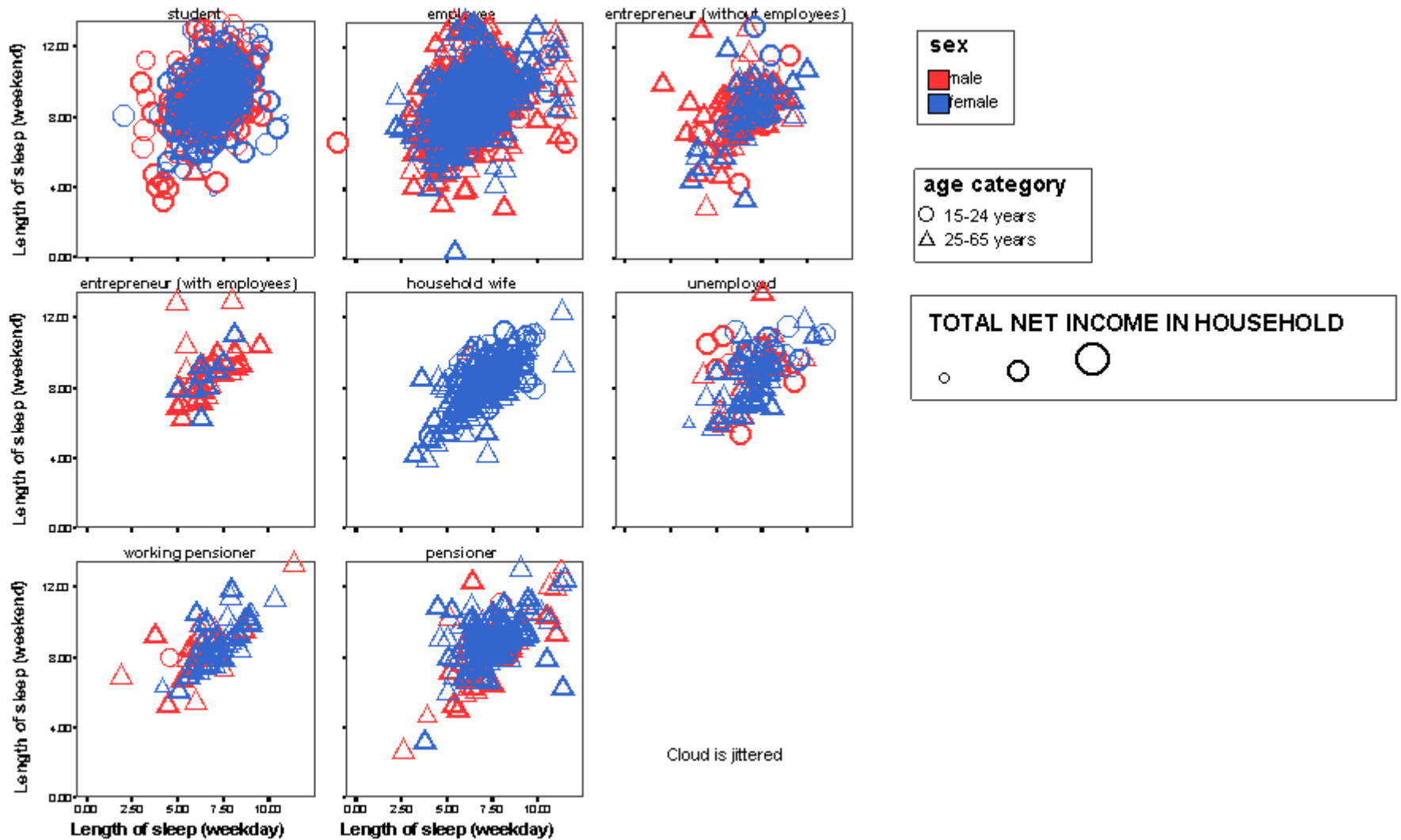
“Tableau-graphique” showing transportation of commercial traffic by variable width (distance), divided bars (height ~ amount), area ~ cost of transport [An early form of the mosaic plot.] Charles Joseph Minard (1781-1870), France



Maximum likelihood estimation depending on two parameters



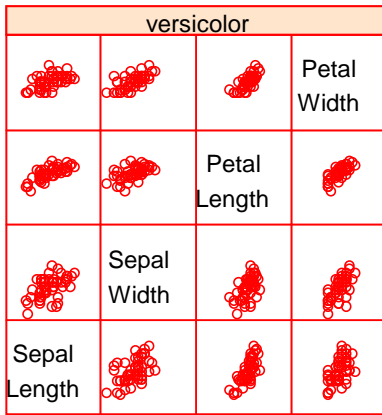
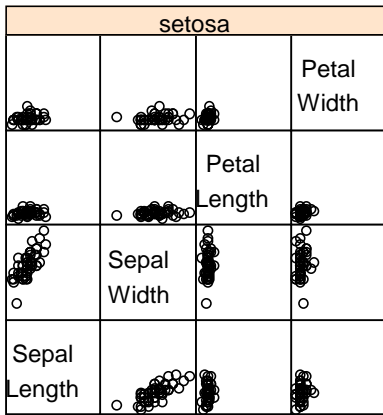
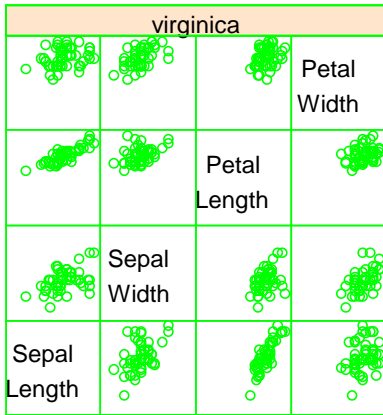
Simulation results
of RCA models



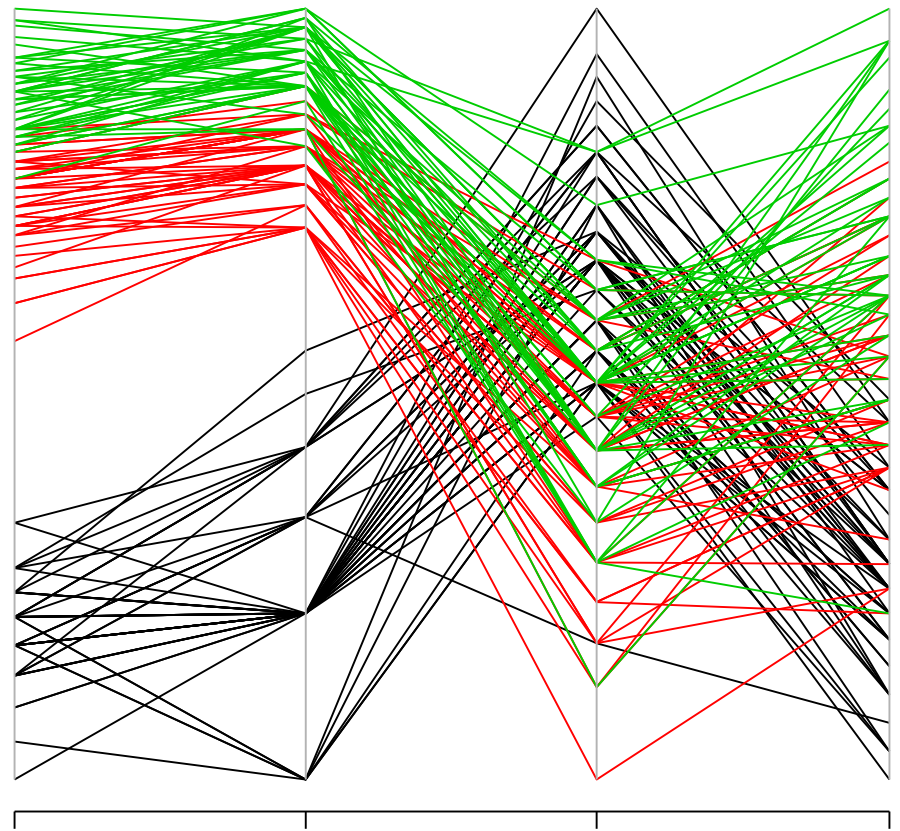
5-dimensional scatter plot



Three Varieties of Iris



Scatter Plot Matrix

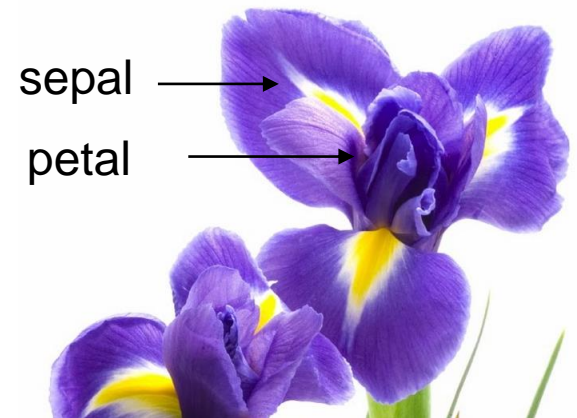


Petal Length

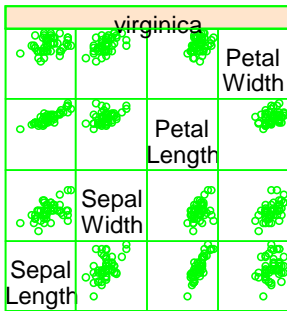
Petal Width

Sepal Width

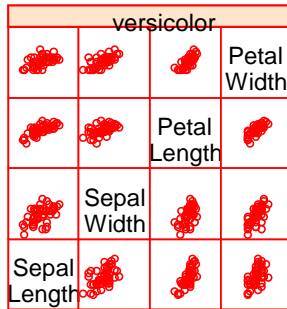
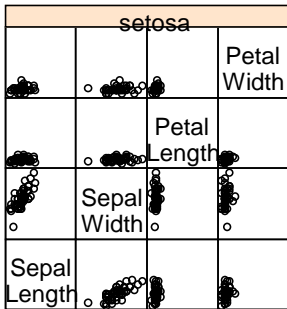
Sepal Length



Scatter plot matrix and parallel plot of iris data

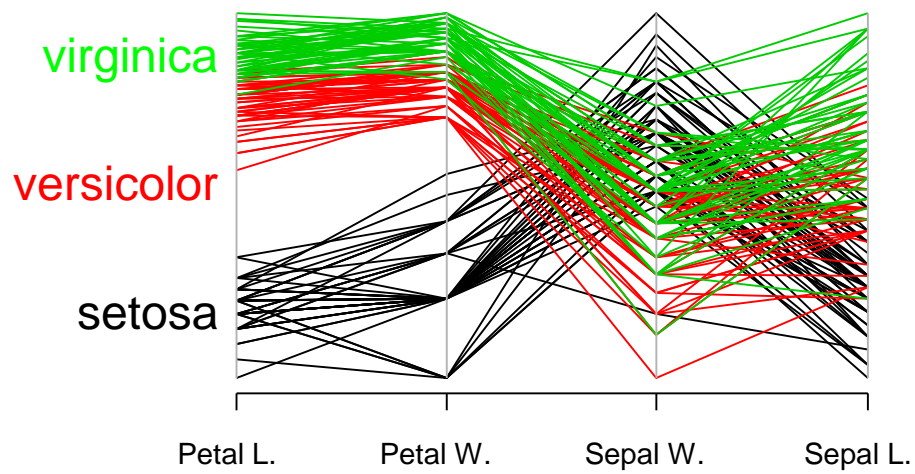


Three Varieties of Iris

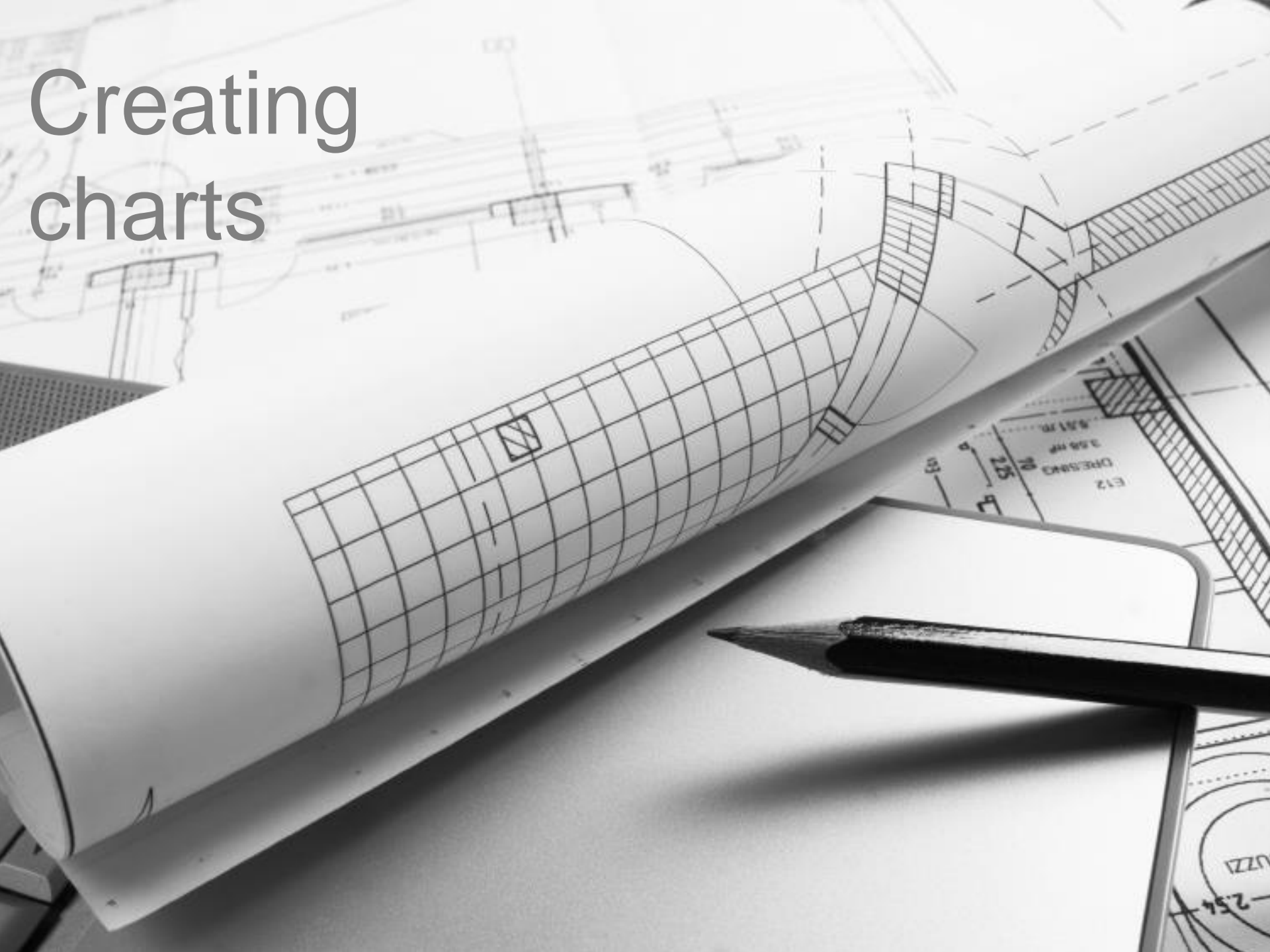


Scatter Plot Matrix





Creating charts



Data
(information)

data preparation
data analysis
statistical modelling

Main message
(knowledge)

comparison selection
chart type selection
chart creation

Chart



Data
(information)

data preparation
data analysis
statistical modelling

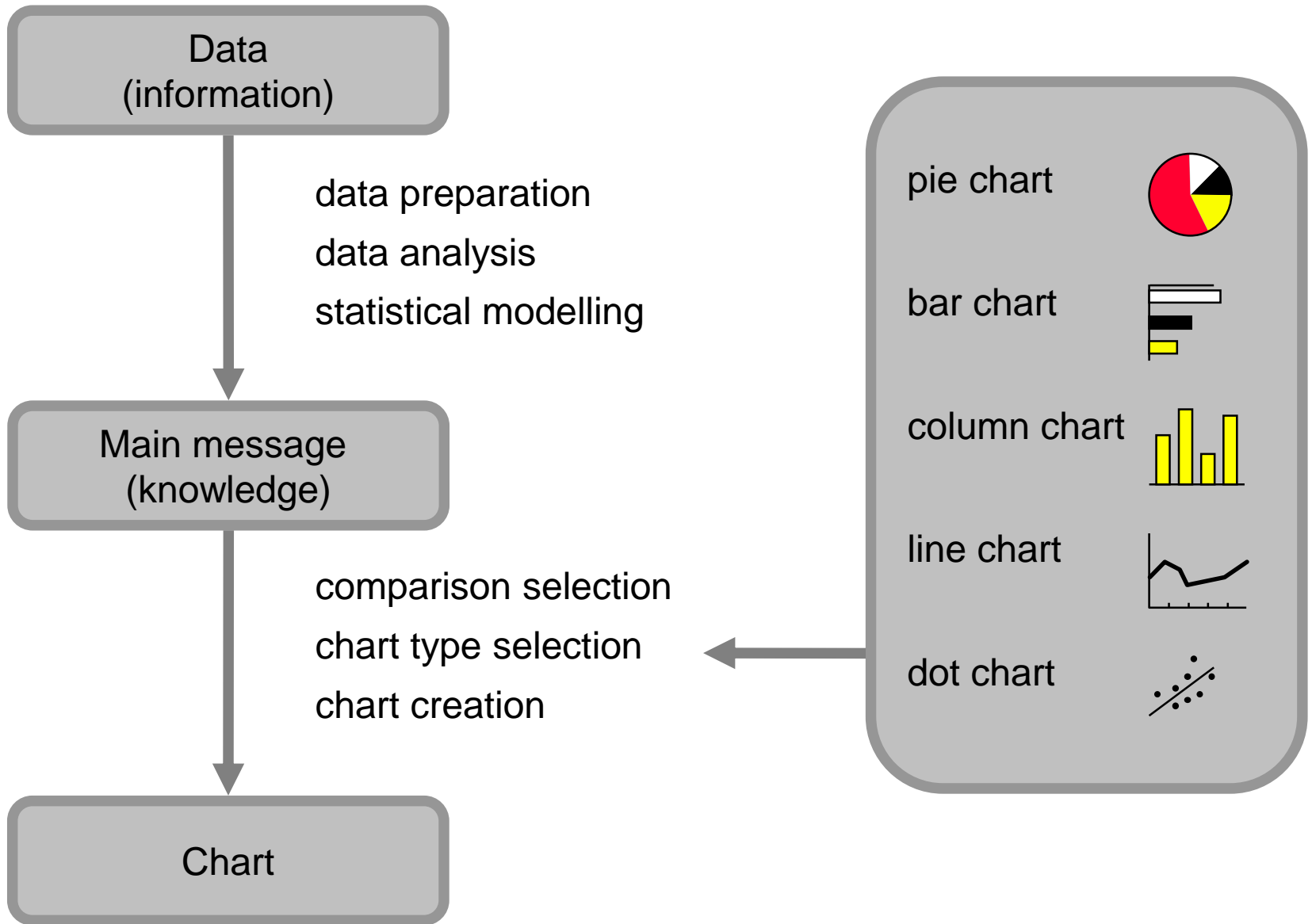
Main message
(knowledge)


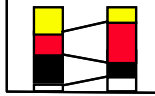

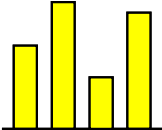
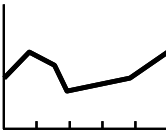
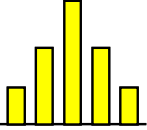

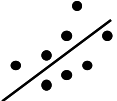
comparison selection
chart type selection
chart creation

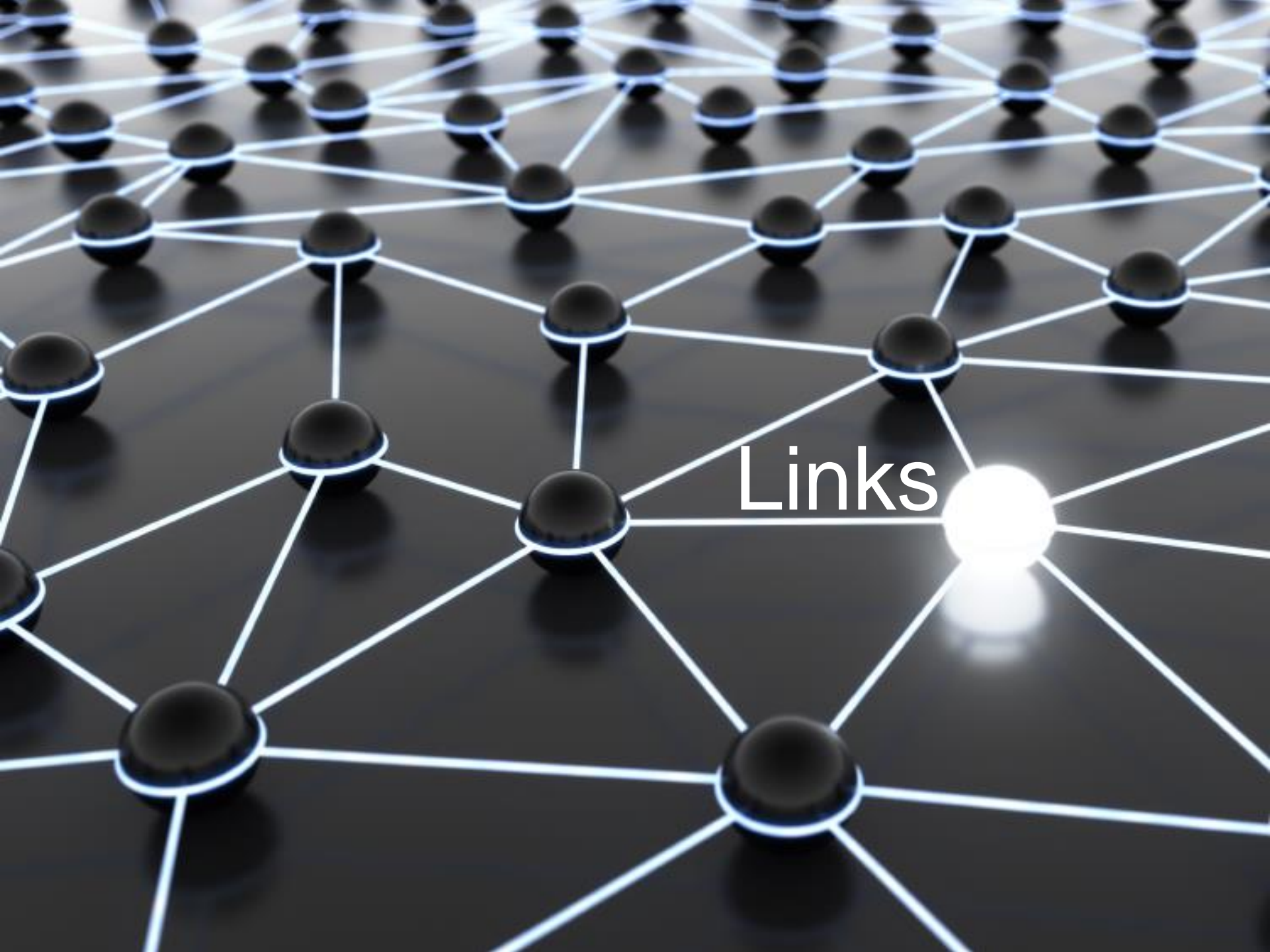
Chart

component
item
time series
frequency distribution
co-relation



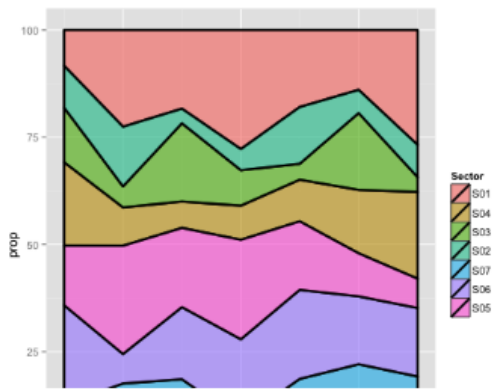


	pie	bar	column	line	dot
component					
item					
time series					
frequency					
co-relation					



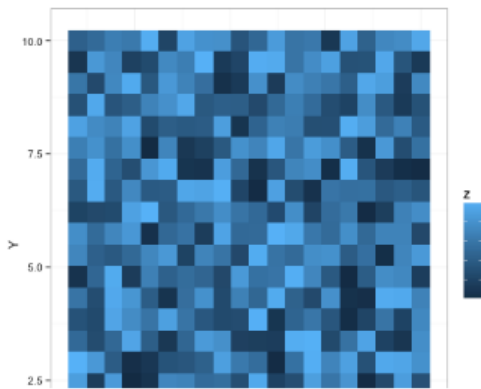
Links

INTERACTIVE



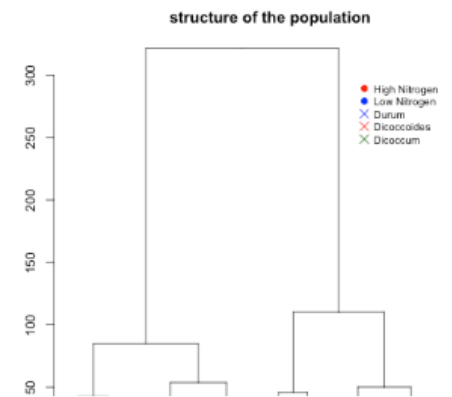
STACKED AREA GRAPH

SCATTER PLOT

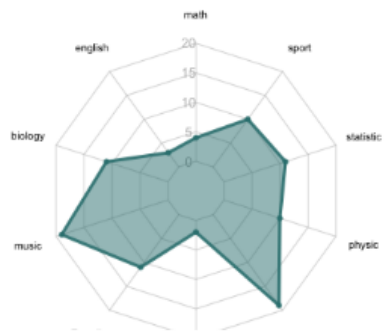


HEATMAP

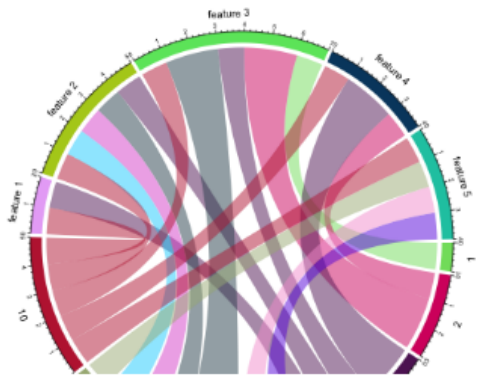
STREAMGRAPH



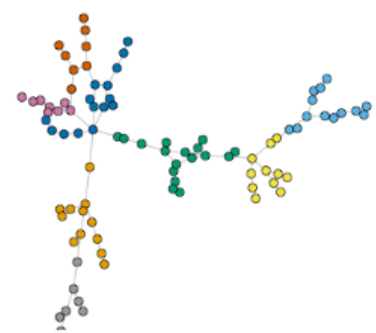
DENDROGRAM



SPIDER / RADAR CHART



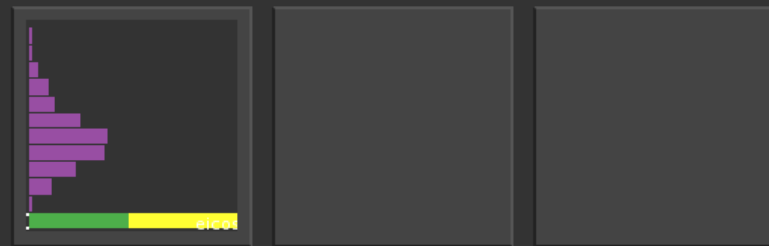
CIRCULAR PLOT



NETWORK

GGobi

Data visualization system



News: [Hack-at-it 2010](#)

Download GGobi for [Windows](#), [Mac](#) and [Linux](#)

Introduction

GGobi is an open source visualization program for exploring high-dimensional data. It provides highly dynamic and interactive graphics such as [tours](#), as well as familiar graphics such as the scatterplot, barchart and parallel coordinates plots. Plots are interactive and linked with [brushing](#) and identification.

GGobi is fully documented in the GGobi book: "[Interactive and Dynamic Graphics for Data Analysis](#)".

If you are interested in how GGobi came to be, you can read more about it on [our history page](#).

NIST SEMATECH

HANDBOOK CHAPTERS

- 1. Explore
- 2. Measure
- 3. Characterize
- 4. Model
- 5. Improve
- 6. Monitor
- 7. Compare
- 8. Reliability

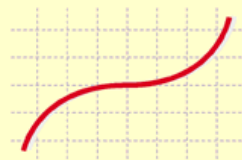
HOW TO USE HANDBOOK

TOOLS & AIDS

SEARCH HANDBOOK

DETAILED CONTENTS

ACKNOWLEDGMENTS



ENGINEERING STATISTICS

H A N D B O O K

Welcome! The goal of this handbook is to help scientists and engineers incorporate statistical methods in their work as efficiently as possible.

To reference the Handbook please use a citation of the form:

NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, date.
(Links to specific pages can also be referenced this way, if suitable.)

A [significant update](#) was made to the Handbook April, 2012

Printer friendly versions of each chapter in the Handbook can be found [here](#).

Feedback on the Handbook sent to handbook@nist.gov is also much appreciated.

[Privacy Policy/Security Notice](#)

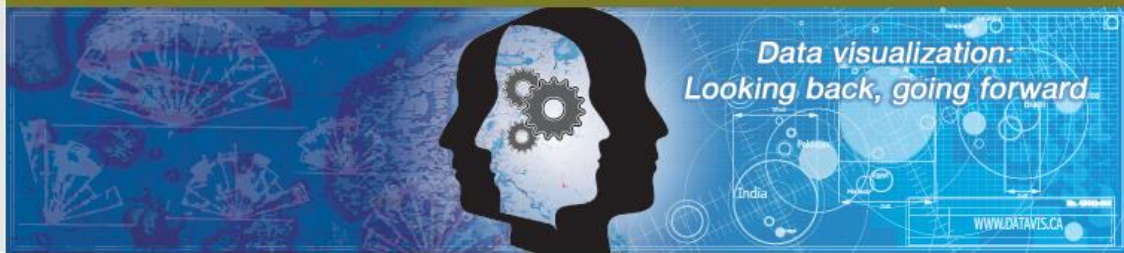
[Disclaimer](#) | [FOIA](#)

DataVis.ca

Michael Friendly
York University



Google™ Custom Search

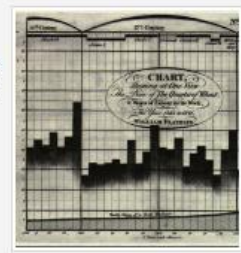


- Home: [datavis.ca](#)
- ▶ [Milestones Project](#)
- ▶ [Data Visualization Gallery](#)
- ▶ [Books](#)
- ▶ [Courses](#)
- ▶ [Online Papers](#)
- ▶ [R Software](#)
- ▶ [SAS Software](#)
- ▶ [Online Applications \(legacy\)](#)
- Personal Info

Milestones Project

The Milestones Project is a comprehensive, visual compendium of significant events in the histories of data visualization, statistical graphics and thematic cartography. This new version features an interactive timeline, an interactive map of authors' birthplaces, and a calendar of significant events in this history.

Quick links: [Main page](#) | [Author map](#) | [Milestones calendar](#)



Data Visualization Gallery

This Gallery of Data Visualization displays some examples of the Best and Worst of Statistical Graphics, with the view that the contrast may be useful, inform current practice, and provide some pointers to both historical and current work.

[Visit Site](#)



Garr Reynolds

Best-selling Author, Speaker

Introduction

Speaking

Books / DVDs

Presentation Tips

Resources

Presentation Zen

Contact

日本語

Introduction

About Garr

Garr Reynolds is an internationally author of best-selling books including *Presentation Zen Design*, and the communication takes the principles reveal simple, concrete tips for communication approach has inspired millions to communicate visually. A sought-after speaker and many in the Fortune 500. An award currently holds the position of Professor at Gaidai University. Garr is a former at Apple, Inc. in Silicon Valley. SA Garr is one of the "100 most-influential of the Zen arts and resident of Japan with his wife, two children, and two



Copyrighted Material
FOREWORD BY GUY KAWASAKI

presentationzen

Simple Ideas on Presentation Design and Delivery



Garr Reynolds

New Riders

Copyrighted Material

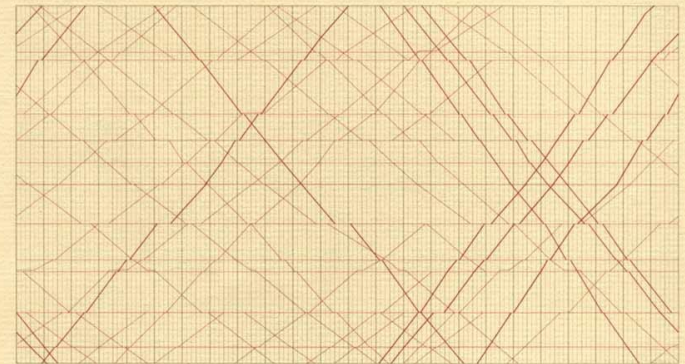
VOICES THAT MATTER™

C. Chen
W. Härdle
A. Unwin
(Editors)

Springer
Handbooks of
Computational
Statistics

Handbook of Data Visualization

 Springer



The Visual Display of Quantitative Information

EDWARD R. TUFTE



Show Me

Data Analytics

data (tracking_TRANSAKCE)

Dimensions

- PROCESS DATETIME (Hours)
- AMOUNT CATEGORIZED
- COUNTRY, CITY
- COUNTRY
- CITY
- HOLDER
- LOC_GEO_LATITUDE
- LOC_GEO_LONGITUDE
- MERCHANT
- MERCHANT_DESC
- MERCHANT_UNIFIED_DESC
- PROCESS DATE
- PROCESS DATETIME
- PROCESS DATETIME (Hours) (group)
- PROCESS DATETIME (Weekdays)
- PROCESS DATETIME (Weekdays) (group)
- TRANSACTION TYPE
- Measure Names

Measures

- AMOUNT
- Latitude (generated)
- Longitude (generated)
- Number of Records
- Measure Values

Pages

Filters

Marks

Automatic

Color Size Text

Detail Tooltip

Columns

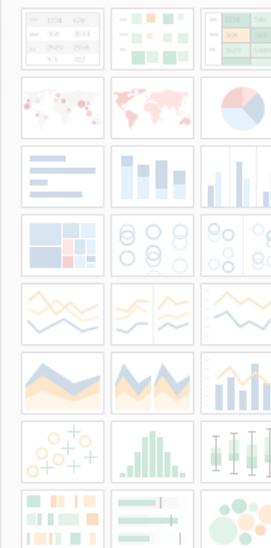
Rows

Sheet 4

Drop field here

Drop field here

Drop field here

**Select or drag data**

Use the Shift or Ctrl key to select multiple fields

Data Source

Sheet 4



