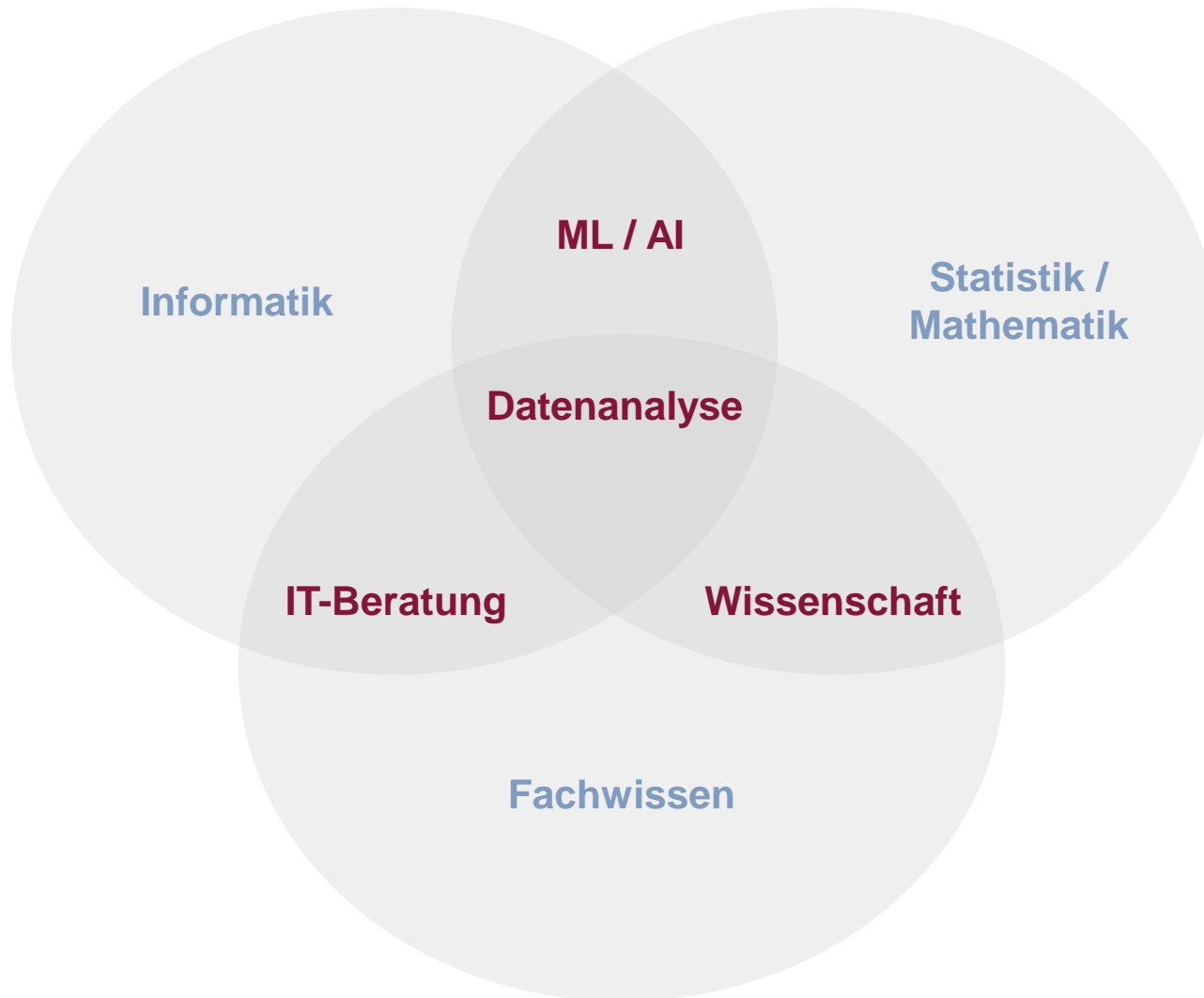


# Datenanalyse

.consulting .solutions .partnership



# Datenanalyse nach Drew Conway



# US Census Income Data

Explorative Datenanalyse

adult.data														
1	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2			
2	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White					
3	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0		
4	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0		
5	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0			
6	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0			
7	49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female				
8	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White					
9	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	1			
10	42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male				
11	37	Private	280464	Some-college	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male				
12	30	State-gov	141297	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	Asian-Pac-Is					

- Erste 12 Instanzen mit
- 15 Variablen

CensusData.R x rawData x

32,561 observations of 15 variables

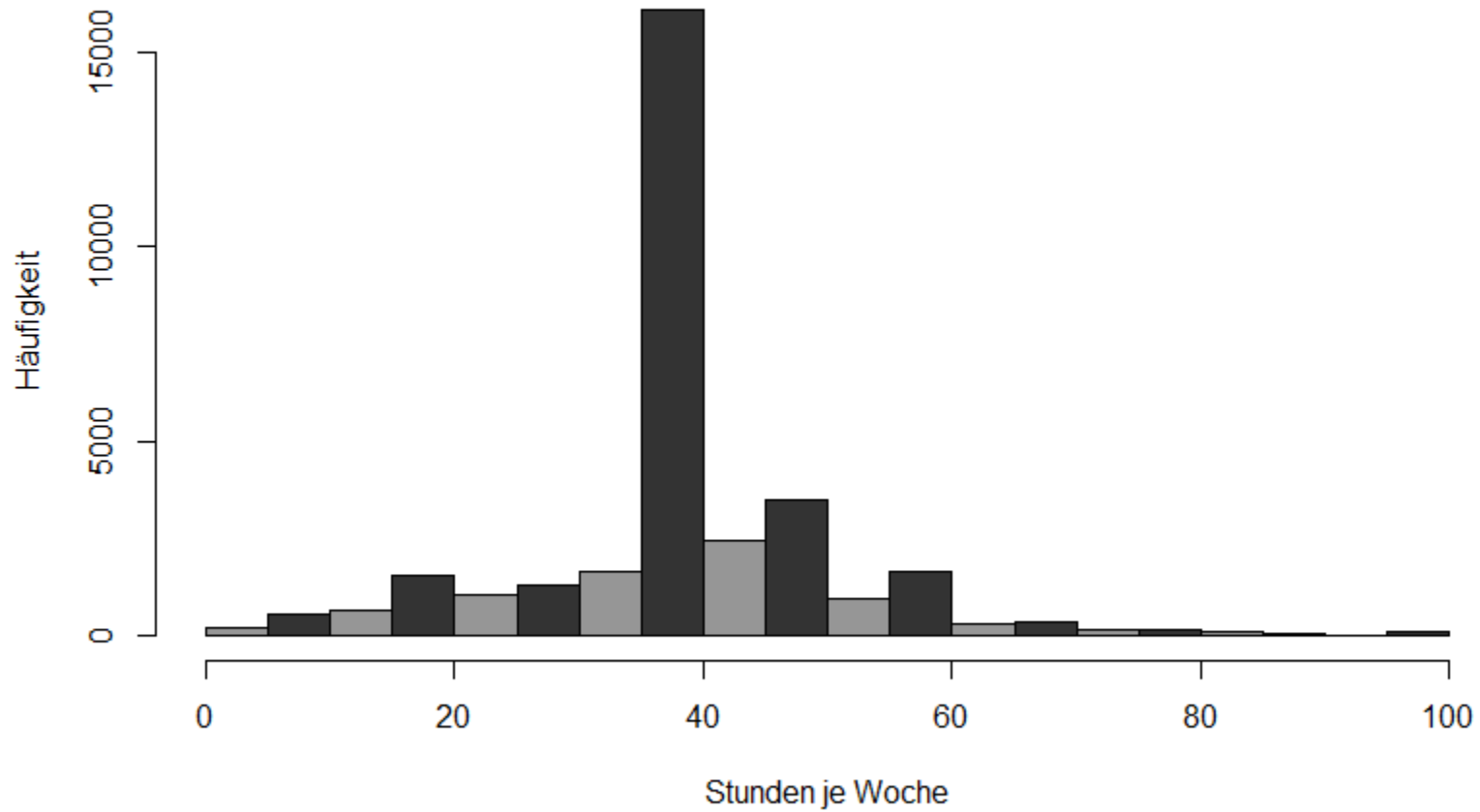
	id	employerKind	fnlwgt	degree	yearsOfEd	maritalStatus	occupation	relationshipRole
1	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family
2	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband
3	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family
4	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband
5	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife
6	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife
7	49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family
8	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband
9	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family
10	42	Private	159449	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband
11	37	Private	280464	Some-college	10	Married-civ-spouse	Exec-managerial	Husband
12	30	State-gov	141297	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband

- Erste 12 Instanzen als Data-Frame mit
- 8 von 15 Variablen

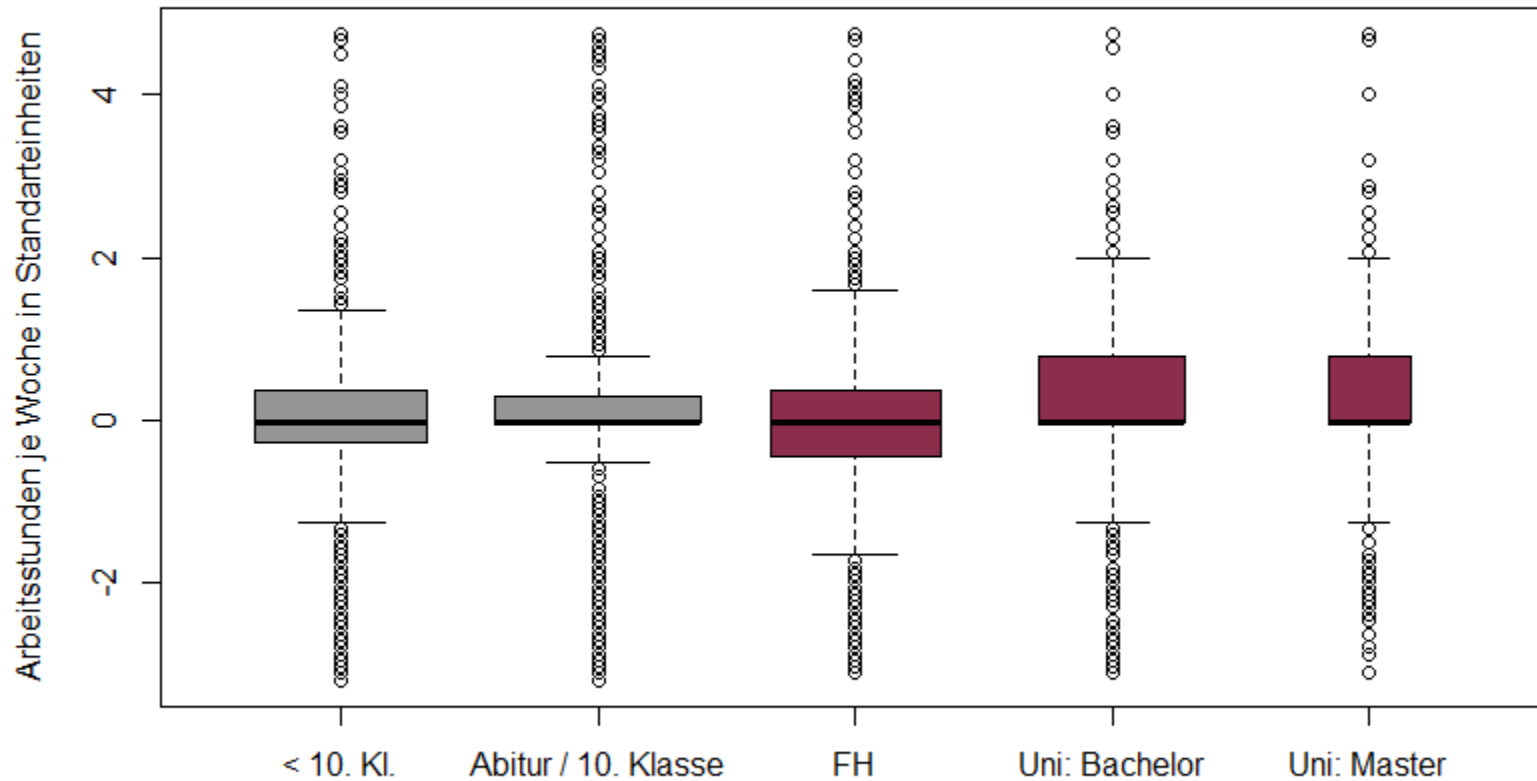
CensusData.R * rawData * 32,561 observations of 20 variables					
incomeGroup	academicLvl	incomeMoreThan50K	capitalDeviation	yearsOfEdStdUnits	workingHoursAWeekStdUnits
<=50K	Bachelor	TRUE	2.7562412	1.13472134	-0.03542890
<=50K	Bachelor	TRUE	-0.3071748	1.13472134	-2.22211900
<=50K	Highschool	TRUE	-0.3071748	-0.42005317	-0.03542890
<=50K	1	TRUE	-0.3071748	-1.19744043	-0.03542890
<=50K	Bachelor	TRUE	-0.3071748	1.13472134	-0.03542890
<=50K	Master	TRUE	-0.3071748	1.52341497	-0.03542890
<=50K	1	TRUE	-0.3071748	-1.97482769	-1.97915343
>50K	Highschool	FALSE	-0.3071748	-0.42005317	0.36951371
>50K	Master	FALSE	3.5009221	1.52341497	0.77445632
>50K	Bachelor	FALSE	3.1020895	1.13472134	-0.03542890
>50K	College	FALSE	-0.3071748	-0.03135955	3.20411198
>50K	Bachelor	FALSE	-0.3071748	1.13472134	-0.03542890

## 5 Sekundärvariablen

Arbeitsstunden je Woche

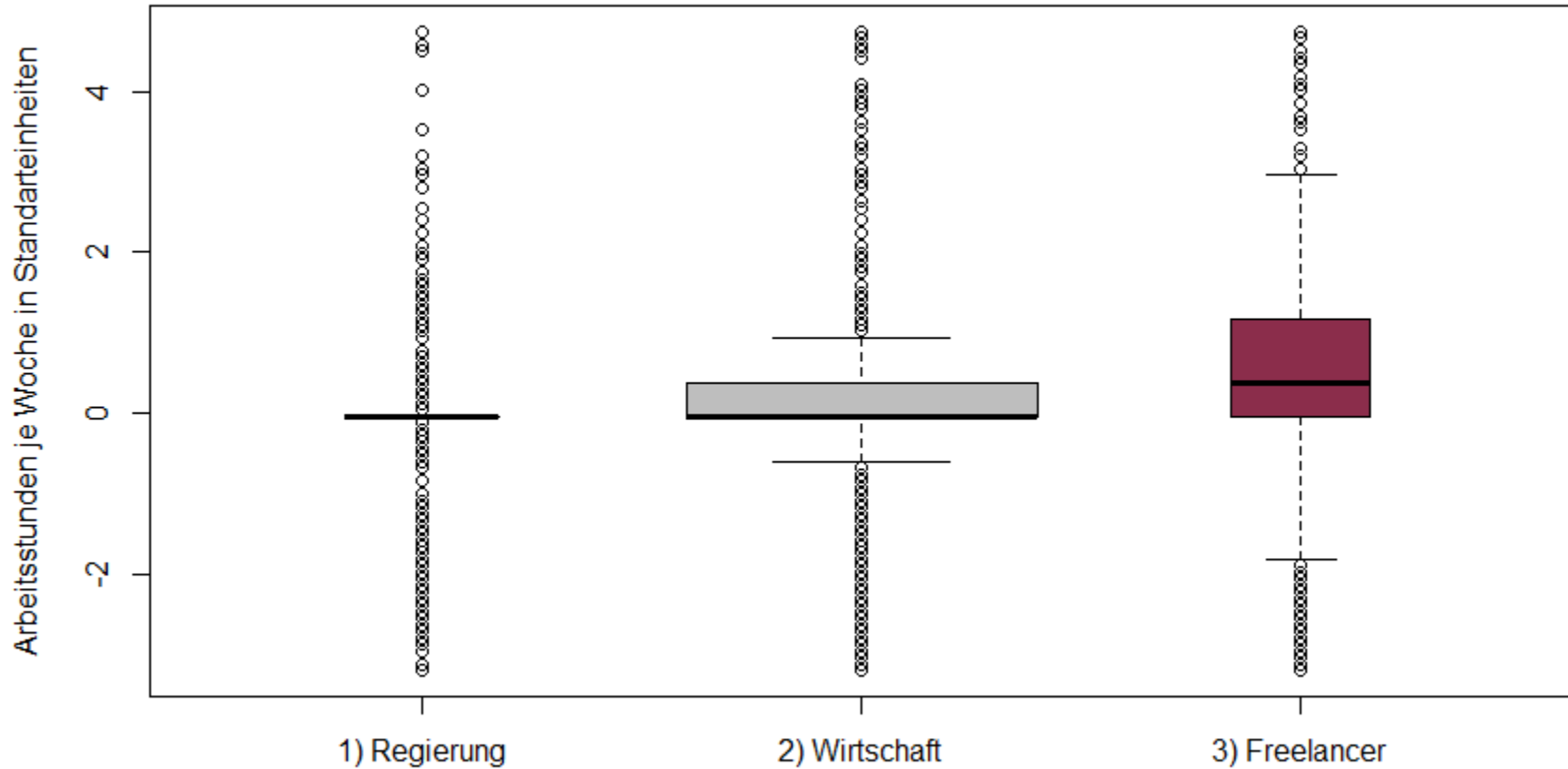


Arbeitsstunden nach Bildungsgrad





## Arbeitsstunden nach Anstellungsart



The screenshot displays the RStudio interface with three main panels:

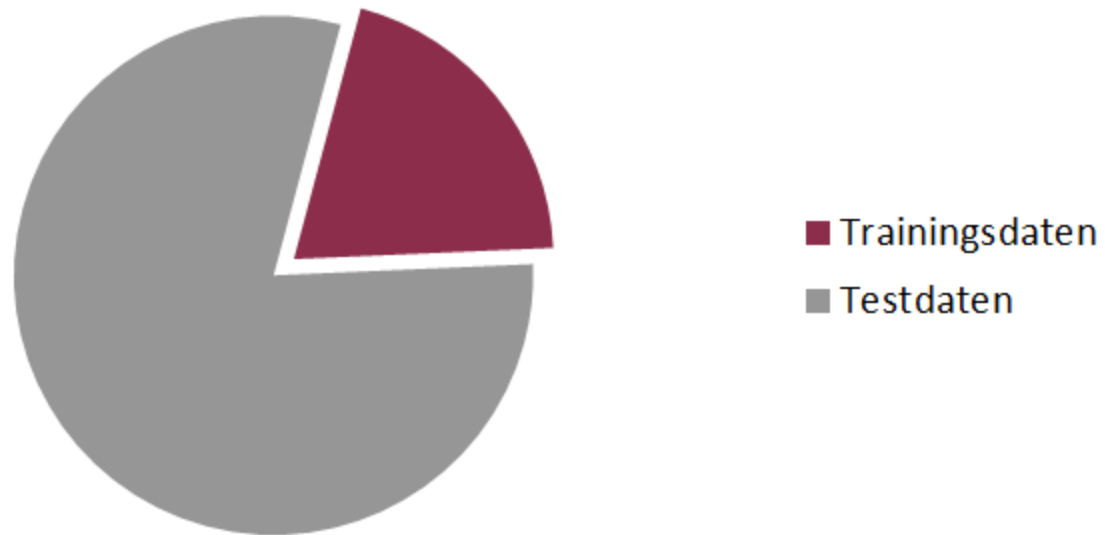
- Script (Source Editor):** Contains R code for setting up a workspace, loading data, and formatting it. The code includes comments and function calls like `library(e1071)`, `set.seed(4711)`, `rm(list = ls()[! (ls() %in% PERSISTENT_CONSTANTS)])`, and `read.csv2`.
- Workspace:** Shows the current environment with variables `PERSISTENT_CONSTANTS` (character [2]) and `WD` (a file path).
- Console:** Displays the execution output of the script, showing the same commands as the script panel. It ends with an error message: `Fehler in setwd(WD) : kann Arbeitsverzeichnis nicht wechseln`.

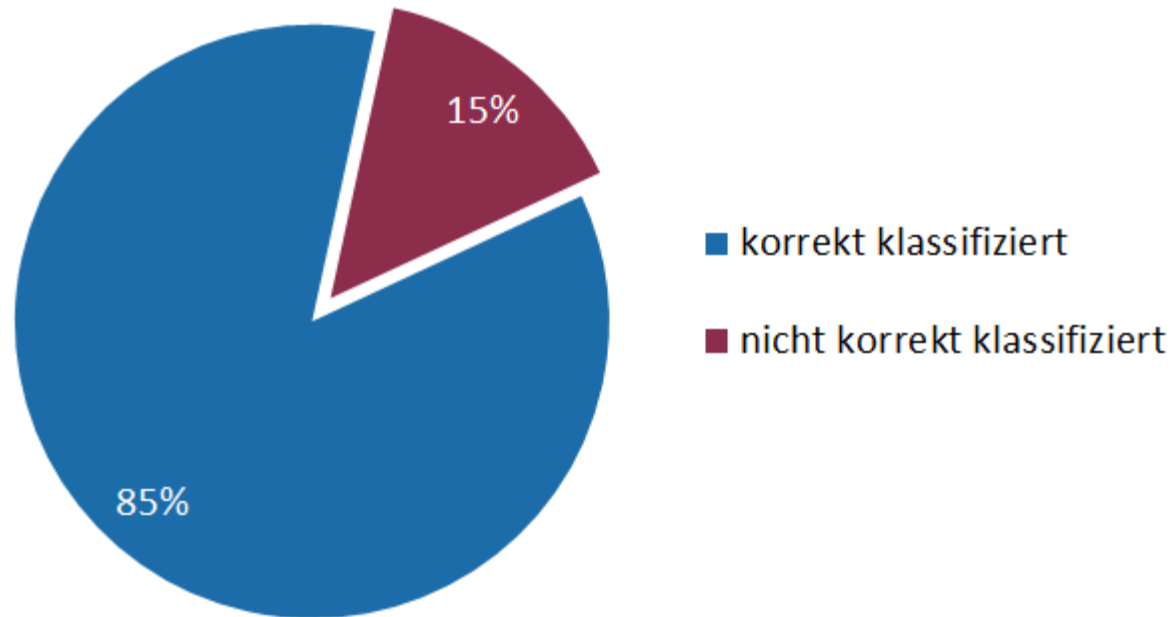
Labels are placed over the panels to identify them:

- Skript** (Script) is placed over the Source Editor.
- Variablen** (Variables) is placed over the Workspace panel.
- Console** is placed over the Console panel.
- Diverse** (Diverse) is placed over the bottom right area of the interface.

# US Census Income Data

Klassifikation



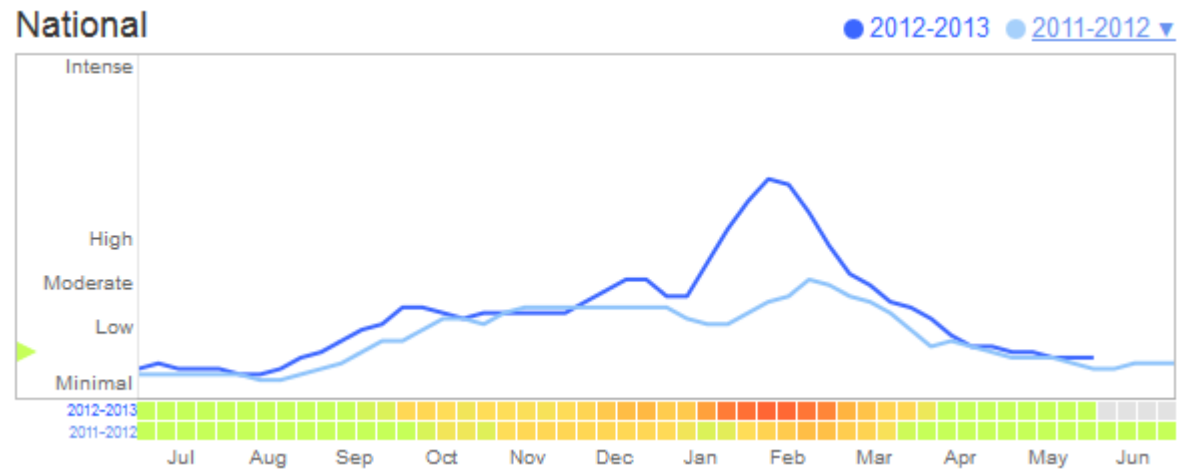


# Anwendungsbeispiele

Google



Google







## LearningR

Training Dataset for R Beginners

Last updated 3 minutes ago



## AddHealth-Data-Analysis

The analysis of biases and influencer in attendance of religious services

Last updated a month ago

[github.com/danielschulz/LearningR](https://github.com/danielschulz/LearningR)



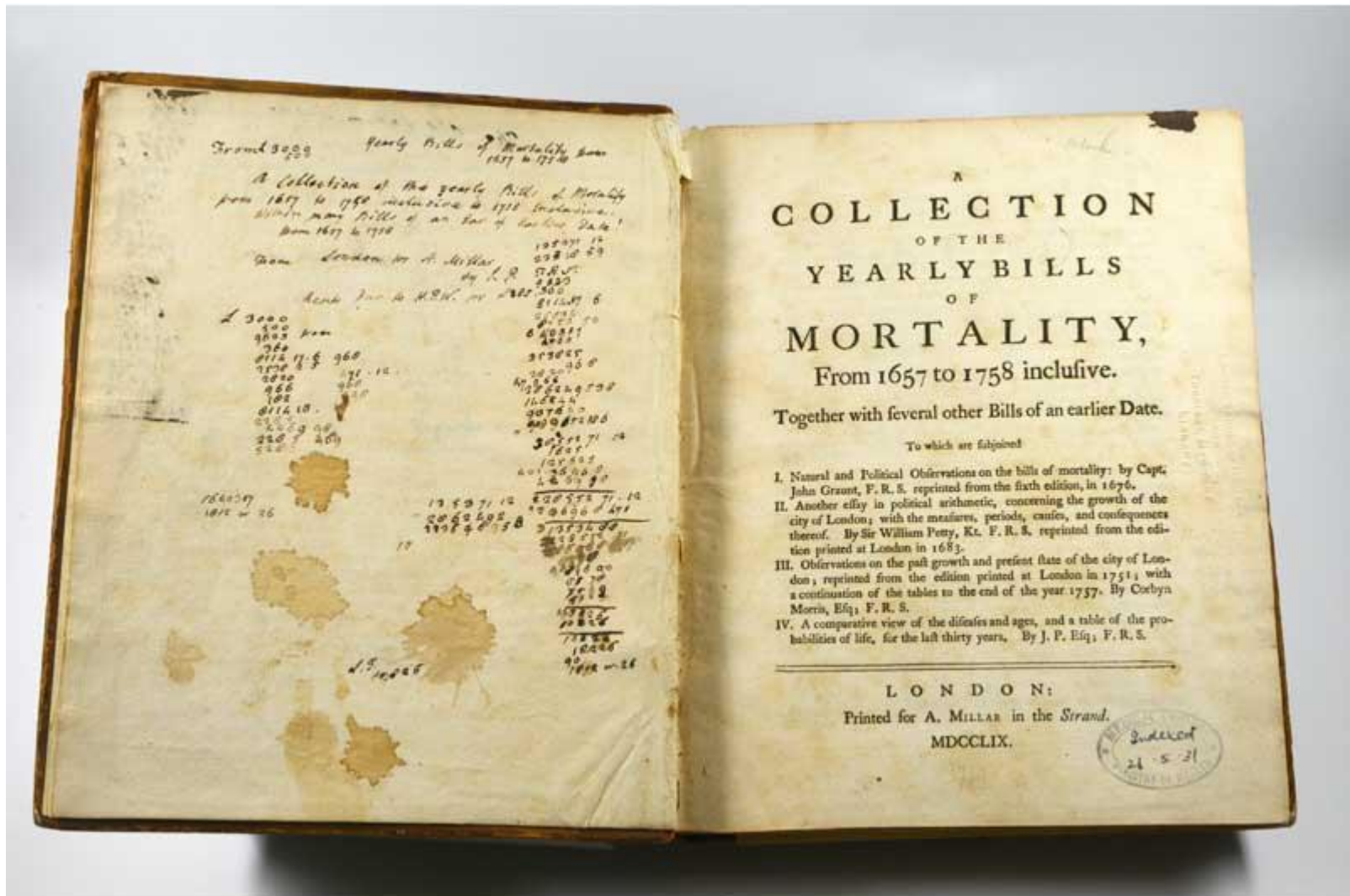
UCI Machine Learning Repository  
[archive.ics.uci.edu/ml](http://archive.ics.uci.edu/ml)

The Kaggle logo, consisting of the word "kaggle" in a blue, lowercase, sans-serif font, with a small trademark symbol (TM) to the upper right of the "e".

[kaggle.com](https://kaggle.com)



[data.gov](http://data.gov)



# Bills of Mortality

# The Table of CASUALTIES.

The Years of our Lord	1647	1648	1649	1650	1651	1652	1653	1654	1655	1656	1657	1658	1659	1660	1661	1662	1663	1664	1665	1666	1667	1668	1669	1670	1671	1672	1673	1674	1675	1676	1677	1678	1679	1680	1681	1682	1683	1684	1685	1686	1687	1688	1689	1690	1691	1692	1693	1694	1695	1696	1697	1698	1699	In Year
Age and Still-born	335	329	327	354	380	381	384	433	483	419	461	467	431	544	499	439	410	445	500	475	50	523	1793	2005	1841	1587	1812	1347	855																									
Small Fever	916	835	889	890	780	834	864	974	743	892	869	1176	909	1095	579	712	661	671	704	623	794	714	2475	2814	3110	1432	1680	1377	1575																									
Stomach and Suddenly	1260	884	751	970	1038	1212	282	1371	689	875	999	1800	303	2148	950	1091	1115	1108	953	1279	1621	1360	4418	6213	1863	4501	4361	4010	2278																									
Stomach	68	74	64	74	100	111	118	80	91	102	113	118	91	67	22	36		17	24	35	20		75	85	280	421	445	177	130																									
Stomach	4	1		3	7	2			4		5	5	3	8	13	8	10	13	6	4		4	54	24	5	12	14	10	9																									
Stomach	3	2	5	1	3	4	3	2	7	3	5	4	7	2	5	2	5	4	4	3		4	14	7	11	10	19	17	61																									
Stomach, Scouring and Flux	153	170	802	389	833	762	200	386	166	366	362	433	346	251	449	438	350	345	273	512	344	330	1567	1400	1432	1811	1161	1597	7818																									
Stomach and Scalded	3	0	10	5	11	6	5	7	10	5	7	4	6	6	3	10	7	5	1	3	12	3	25	19	24	31	20	19	135																									
Stomach	1			1		2	1	1														1	3	4	2	4	3	38																										
Stomach, Gangrene and Fistula	26	29	31	25	31	53	30	37	73	31	24	35	63	52	20	14	23	28	27	30	24	30	85	112	103	157	150	124	609																									
Stomach, Sore-mouth and Throat	66	28	54	41	65	51	53	72	44	81	19	27	73	68	6	4	4	1				1	74	15	70	190	244	161	113	659																								
Stomach	161	106	114	117	200	213	158	192	177	201	210	213	225	194	150	157	112	171	132	143	163	210	590	600	498	769	859	490	3364																									
Stomach and Infants	1369	1254	1065	950	1237	1280	1050	1343	1089	1393	1162	1144	858	1121	259	2174	1035	2268	2130	2315	2113	1895	9277	8453	4078	1910	1788	4519	22106																									
Stomach and Wind	103	71	85	82	76	101	80	101	85	120	113	117	116	167	48	57						37	50	105	87	141	359	497	247	1350																								
Stomach and Cough	1423	2200	1388	1588	2350	2410	2280	1868	2000	3184	2757	1610	1902	3414	1827	1910	1713	1797	1754	1855	2080	1477	5157	8260	5990	9914	1215	7197	44487																									
Stomach and Cough	684	491	530	493	569	655	606	828	701	1027	807	841	742	1031	52	87	18	241	221	386	418	709	498	1734	1198	1650	1177	1314	9073																									
Stomach																						1	0	0	0	0	0	0	1	2																								
Stomach																						1	5	2	2	3	10	0	4	13	47	18																						
Stomach, Sore and Tympany	185	434	421	506	444	556	617	704	660	706	631	631	646	672	235	252	279	280	260	250	315	369	1048	1718	1518	231	1292	1304	8623																									
Stomach, Sore and Tympany	47	40	30	27	49	50	53	30	43	49	63	60	57	48	43	33	29	34	37	21	32	45	139	147	144	182	215	130	827																									
Stomach, Sore and Tympany																						1	4	2	4	2	2	2	2	2																								
Stomach, Sore and Tympany	8	17	22	43	24	12	18	21	19	22	20	15	7	18	19	13	11	18	13	13	11	13	02	51	97	70	79	51	184																									
Stomach, Sore and Tympany																						1	27	24	10	8	8	5	74																									
Stomach, Sore and Tympany	3	2	2	3		3	4	1	4	3	1	4	5	3	10	7	7	3	5	4		127	701	1340	1913	2753	3161	2781	10570																									
Stomach, Sore and Tympany	139	400	1190	184	535	1279	139	812	1294	823	835	402	1523	354	72	40	58	531	70	1354	191	14	81	69	29	34	2	29	243																									
Stomach, Sore and Tympany	6	6	9	8	7	9	14	4	3	4	5	11	2	8	16	33	20	0	13	8	24	22	33	48	80	81	110	81	192																									
Stomach, Sore and Tympany	15	25	13	15	21	20	20	20	29	23	23	55	51	31	17	12	12	12	7	17	13	3	2	3	10	5	2	2	21																									
Stomach, Sore and Tympany	4	4	1		3	2			1	1												6	14	24	35	25	34	20	134																									
Stomach, Sore and Tympany	5	5	12	5	7	7	5	6	8	7	8	13	14	2	2	5	3	4	4	5	7	20	71	50	48	59	43	47	279																									
Stomach, Sore and Tympany	12	13	14	7	17	14	11	17	10	13	10	12	13	4	18	20	22	11	14	17	1	7	37	13	48	47	7	3	223																									
Stomach, Sore and Tympany	11	10	11	14	9	14	15	9	14	16	24	18	11	30	8	8	6	15				3	8	2	0	14	14	12	40	611																								
Stomach, Sore and Tympany	1	11	2		2	0	6	5	3	4	5	31	20									63	184	197	180	212	223	180	590																									
Stomach, Sore and Tympany	57	21	20	40	41	43	57	71	61	41	40	77	102	76	47	59	35	42	10	43	54	11	47	15	00	4	6	10	00																									

Wöchentliche Todesstatistiken

# Resumée

- Google's Chef-Ökonom Hal Varian
  - „The next sexy job“
  - „The ability to take data – to be able to understand it, to process it, to extract value from it, to communicate it – that's going to be a hugely important skill.“
  - New York Times, 2009
- „Hot new gig in tech“ – Fortune





**Vielen Dank für Ihre Aufmerksamkeit**



.consulting .solutions .partnership

