

Datenformen lang und breit & Daten zusammenfügen

rstatsZH - Data Science mit R

Lars Schöbitz

Apr 8, 2025

Lernziele (für diese Woche)

1. Die Lernenden können Funktionen aus dem R-Paket `tidyverse` anwenden, um ihre Daten von einem breiten in ein langes Format und umgekehrt umzuwandeln.
2. Die Lernenden können Funktionen aus dem `dplyr` R-Paket anwenden, um mehrere Datensätze zu verknüpfen.
3. Die Lernenden können die drei grundlegenden Elemente des Tidy Data Konzeptes wiedergeben.

Teil 1: Daten drehen

Ihr seid dran: 01-treibhausgase-ihr.qmd

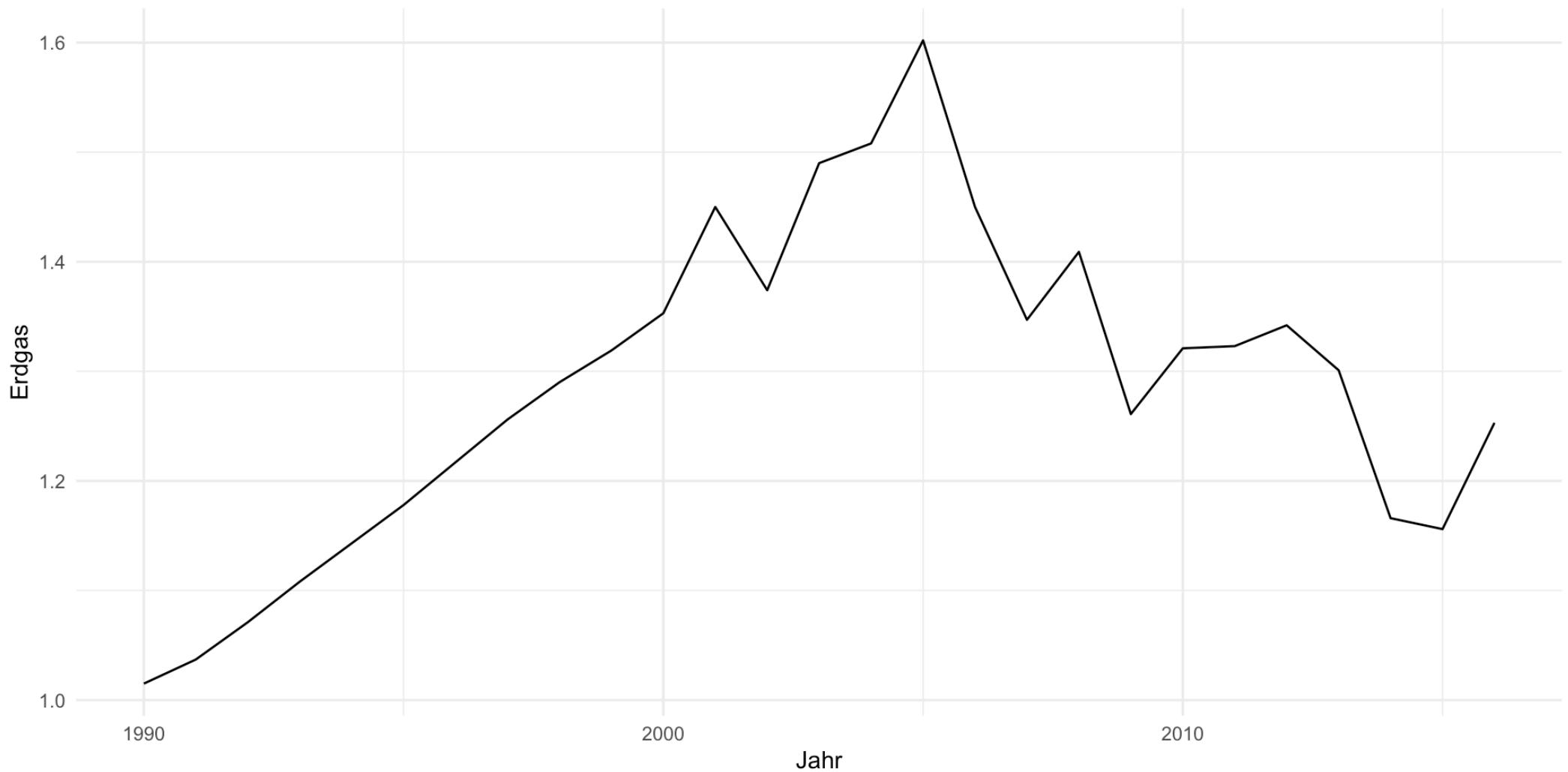
1. Öffne [posit.cloud](#) in deinem Browser (verwende dein Lesezeichen).
2. Öffne den rstatszh-k010 Arbeitsbereich (Workspace) für den Kurs.
3. Klicke auf Start neben md-07-uebungen.
4. Suche im Dateimanager im Fenster unten rechts die Datei 01-treibhausgase-ihr.qmd und klicke darauf, um sie im Fenster oben links zu öffnen.
5. Folge den Anweisungen in der Datei.

Daten-Struktur: Treibhausgase

Jahr	Strom	Kerosin	Diesel	Benzin	Holz_UW_BG_SK	Fernwaerme	Erdgas	Heizoel_EL
1990	0.324	0.638	0.418	1.087	0.021	0.228	1.015	2.445
1991	0.292	0.614	0.415	1.081	0.020	0.227	1.037	2.338
1992	0.263	0.637	0.418	1.083	0.020	0.229	1.071	2.258
1993	0.243	0.651	0.419	1.093	0.020	0.232	1.108	2.185
1994	0.229	0.661	0.416	1.099	0.020	0.234	1.143	2.109
1995	0.242	0.689	0.415	1.102	0.020	0.237	1.178	2.033
1996	0.256	0.706	0.415	1.112	0.021	0.240	1.217	1.963
1997	0.397	0.730	0.417	1.118	0.021	0.243	1.256	1.892
1998	0.539	0.759	0.418	1.115	0.021	0.245	1.290	1.813
1999	0.667	0.803	0.422	1.107	0.020	0.246	1.319	1.728
2000	0.790	0.833	0.432	1.095	0.021	0.248	1.353	1.650
2001	1.091	0.782	0.445	1.080	0.019	0.250	1.450	1.189
2002	1.381	0.719	0.453	1.057	0.019	0.251	1.374	1.498
2003	1.169	0.642	0.470	1.032	0.018	0.286	1.490	1.317
2004	0.954	0.601	0.486	1.003	0.019	0.298	1.508	1.305
2005	0.170	0.605	0.503	0.970	0.020	0.300	1.602	1.281
2006	0.176	0.629	0.504	0.889	0.019	0.255	1.450	1.425
2007	0.176	0.665	0.508	0.809	0.017	0.218	1.347	1.210
2008	0.180	0.701	0.507	0.738	0.018	0.243	1.409	1.155
2009	0.183	0.666	0.516	0.683	0.020	0.227	1.261	1.246

Jahr	Strom	Kerosin	Diesel	Benzin	Holz_UW_BG_SK	Fernwaerme	Erdgas	Heizoel_EL
2010	0.179	0.671	0.519	0.639	0.024	0.250	1.321	1.255
2011	0.180	0.708	0.508	0.621	0.020	0.145	1.323	1.205
2012	0.185	0.717	0.565	0.596	0.028	0.181	1.342	1.252
2013	0.195	0.720	0.590	0.573	0.036	0.222	1.301	1.170
2014	0.346	0.716	0.597	0.545	0.035	0.142	1.166	0.922
2015	0.398	0.732	0.607	0.491	0.038	0.181	1.156	0.867
2016	0.460	0.750	0.651	0.479	0.059	0.197	1.253	0.896

Wie würdest du dieses Diagramm erstellen?

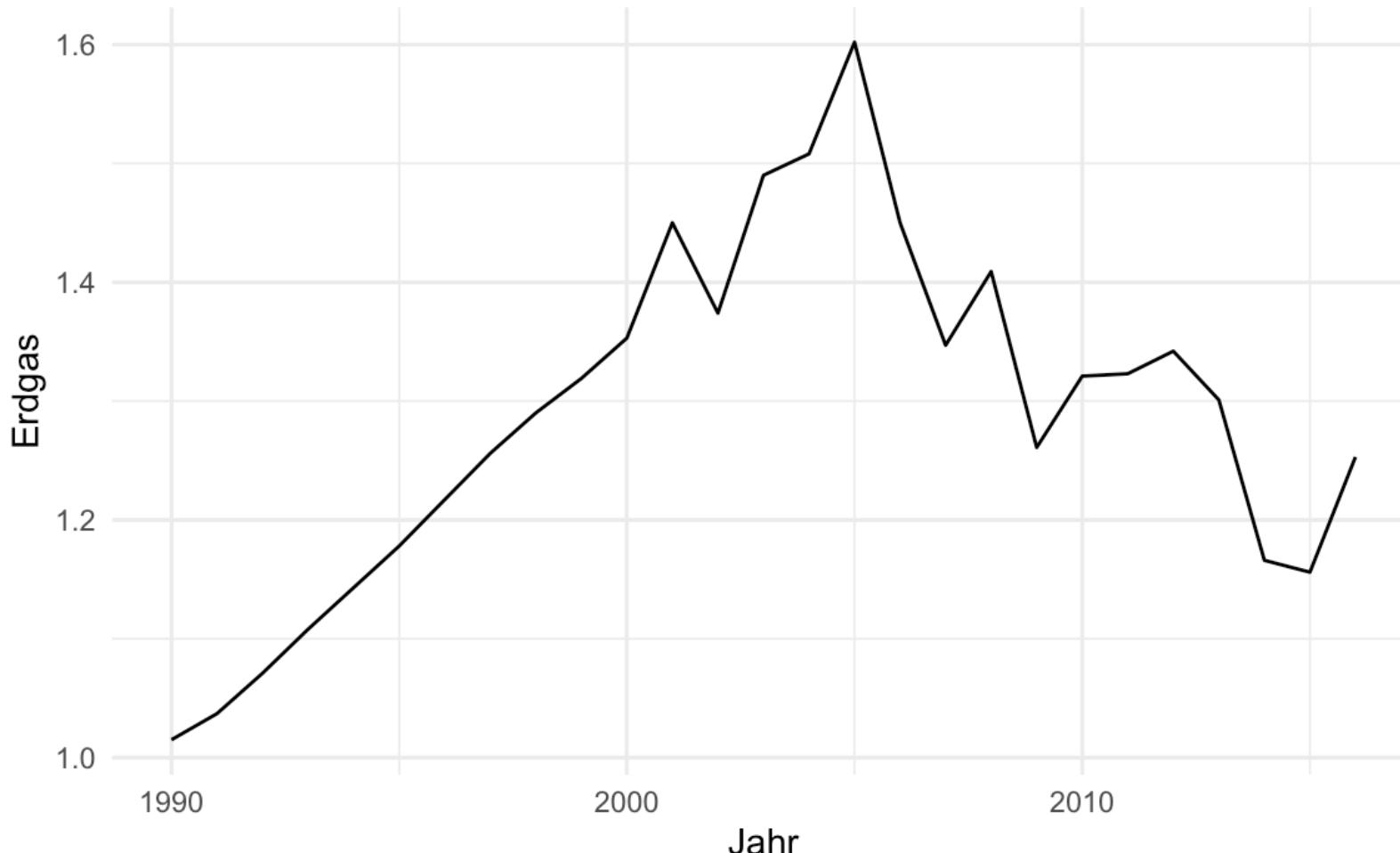


Zwei Variablen

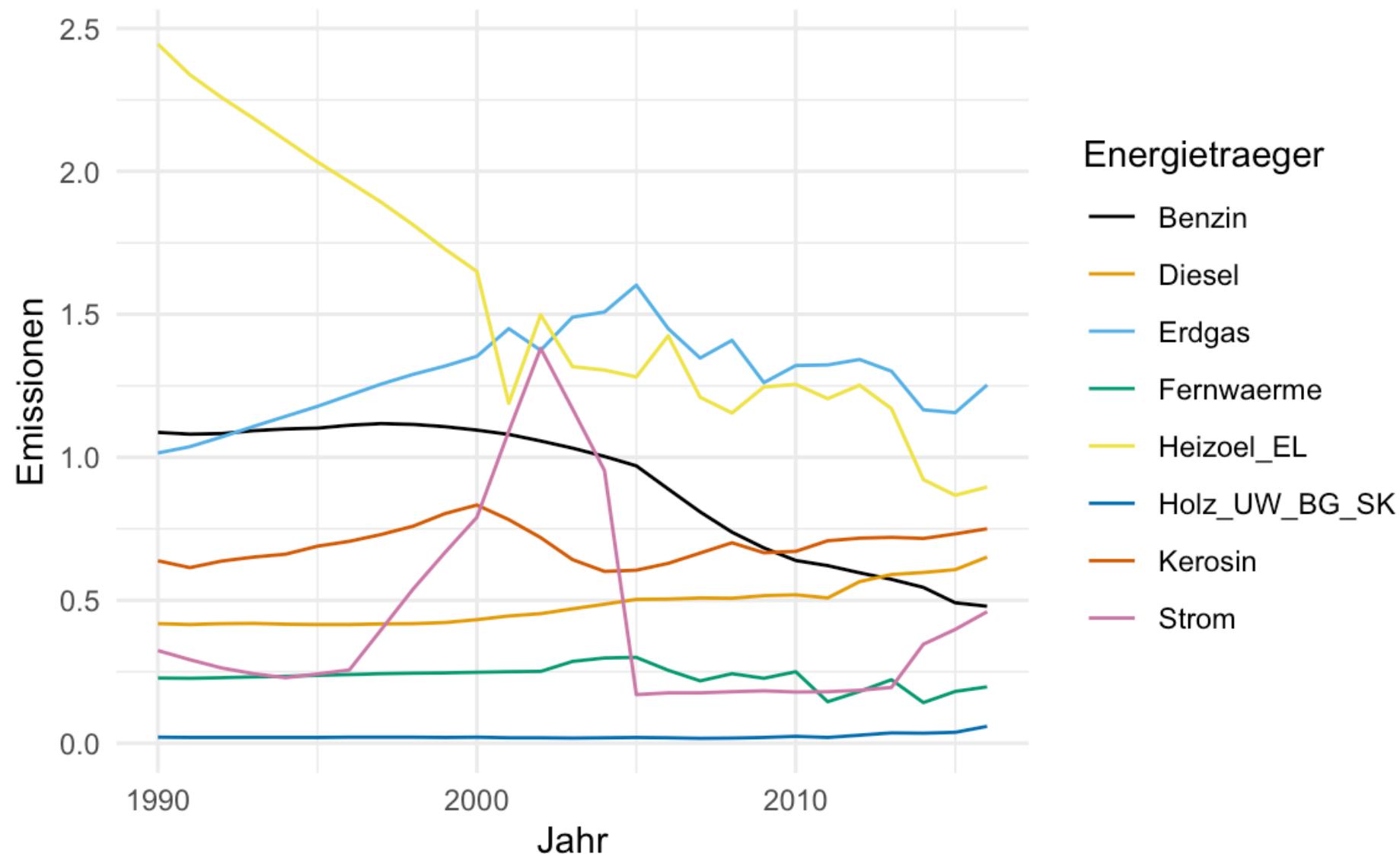
```
# A tibble: 27 × 2
  Jahr Erdgas
  <dbl>   <dbl>
1 1990    1.01
2 1991    1.04
3 1992    1.07
4 1993    1.11
5 1994    1.14
6 1995    1.18
7 1996    1.22
8 1997    1.26
9 1998    1.29
10 1999   1.32
# ... with 17 more rows
```

Zwei Variablen -> Zwei visuelle Eigenschaften (aesthetics)

```
1 ggplot(data = ghg,  
2         mapping = aes(x = Jahr, y = Erdgas)) +  
3         geom_line() +  
4         theme_minimal()
```



Wie würdest du dieses Diagramm erstellen?



Daten sind in einem breitem Format

Jahr	Strom	Kerosin	Diesel	Benzin	Holz_UW_BG_SK	Fernwaerme	Erdgas	Heizoel_EL
1990	0.324	0.638	0.418	1.087	0.021	0.228	1.015	2.445
1991	0.292	0.614	0.415	1.081	0.020	0.227	1.037	2.338
1992	0.263	0.637	0.418	1.083	0.020	0.229	1.071	2.258
1993	0.243	0.651	0.419	1.093	0.020	0.232	1.108	2.185
1994	0.229	0.661	0.416	1.099	0.020	0.234	1.143	2.109
1995	0.242	0.689	0.415	1.102	0.020	0.237	1.178	2.033
1996	0.256	0.706	0.415	1.112	0.021	0.240	1.217	1.963
1997	0.397	0.730	0.417	1.118	0.021	0.243	1.256	1.892
1998	0.539	0.759	0.418	1.115	0.021	0.245	1.290	1.813
1999	0.667	0.803	0.422	1.107	0.020	0.246	1.319	1.728
2000	0.790	0.833	0.432	1.095	0.021	0.248	1.353	1.650
2001	1.091	0.782	0.445	1.080	0.019	0.250	1.450	1.189
2002	1.381	0.719	0.453	1.057	0.019	0.251	1.374	1.498
2003	1.169	0.642	0.470	1.032	0.018	0.286	1.490	1.317
2004	0.954	0.601	0.486	1.003	0.019	0.298	1.508	1.305
2005	0.170	0.605	0.503	0.970	0.020	0.300	1.602	1.281
2006	0.176	0.629	0.504	0.889	0.019	0.255	1.450	1.425
2007	0.176	0.665	0.508	0.809	0.017	0.218	1.347	1.210
2008	0.180	0.701	0.507	0.738	0.018	0.243	1.409	1.155
2009	0.183	0.666	0.516	0.683	0.020	0.227	1.261	1.246

Jahr	Strom	Kerosin	Diesel	Benzin	Holz_UW_BG_SK	Fernwaerme	Erdgas	Heizoel_EL
2010	0.179	0.671	0.519	0.639	0.024	0.250	1.321	1.255
2011	0.180	0.708	0.508	0.621	0.020	0.145	1.323	1.205
2012	0.185	0.717	0.565	0.596	0.028	0.181	1.342	1.252
2013	0.195	0.720	0.590	0.573	0.036	0.222	1.301	1.170
2014	0.346	0.716	0.597	0.545	0.035	0.142	1.166	0.922
2015	0.398	0.732	0.607	0.491	0.038	0.181	1.156	0.867
2016	0.460	0.750	0.651	0.479	0.059	0.197	1.253	0.896

Du brauchst: ein langes Format

Jahr	Energietraeger	Emissionen
1990	Strom	0.324
1990	Kerosin	0.638
1990	Diesel	0.418
1990	Benzin	1.087
1990	Holz_UW_BG_SK	0.021
1990	Fernwaerme	0.228
1990	Erdgas	1.015
1990	Heizoel_EL	2.445
1991	Strom	0.292
1991	Kerosin	0.614
1991	Diesel	0.415
1991	Benzin	1.081
1991	Holz_UW_BG_SK	0.020
1991	<small>@restatehiko18</small> Fernwaerme	0.227

Jahr	Energietraeger	Emissionen
1991	Erdgas	1.037
1991	Heizoel_EL	2.338
1992	Strom	0.263
1992	Kerosin	0.637
1992	Diesel	0.418
1992	Benzin	1.083
1992	Holz_UW_BG_SK	0.020
1992	Fernwaerme	0.229
1992	Erdgas	1.071
1992	Heizoel_EL	2.258
1993	Strom	0.243
1993	Kerosin	0.651
1993	Diesel	0.419
1993	Benzin	1.093
1993	Holz_UW_BG_SK	0.020
1993	Fernwaerme	0.232

Jahr	Energietraeger	Emissionen
1993	Erdgas	1.108
1993	Heizoel_EL	2.185
1994	Strom	0.229
1994	Kerosin	0.661
1994	Diesel	0.416
1994	Benzin	1.099
1994	Holz_UW_BG_SK	0.020
1994	Fernwaerme	0.234
1994	Erdgas	1.143
1994	Heizoel_EL	2.109
1995	Strom	0.242
1995	Kerosin	0.689
1995	Diesel	0.415
1995	Benzin	1.102
1995	Holz_UW_BG_SK	0.020
1995	Fernwaerme	0.237

Jahr	Energietraeger	Emissionen
1995	Erdgas	1.178
1995	Heizoel_EL	2.033
1996	Strom	0.256
1996	Kerosin	0.706
1996	Diesel	0.415
1996	Benzin	1.112
1996	Holz_UW_BG_SK	0.021
1996	Fernwaerme	0.240
1996	Erdgas	1.217
1996	Heizoel_EL	1.963
1997	Strom	0.397
1997	Kerosin	0.730
1997	Diesel	0.417
1997	Benzin	1.118
1997	Holz_UW_BG_SK	0.021
1997	Fernwaerme	0.243

Jahr	Energietraeger	Emissionen
1997	Erdgas	1.256
1997	Heizoel_EL	1.892
1998	Strom	0.539
1998	Kerosin	0.759
1998	Diesel	0.418
1998	Benzin	1.115
1998	Holz_UW_BG_SK	0.021
1998	Fernwaerme	0.245
1998	Erdgas	1.290
1998	Heizoel_EL	1.813
1999	Strom	0.667
1999	Kerosin	0.803
1999	Diesel	0.422
1999	Benzin	1.107
1999	Holz_UW_BG_SK	0.020
1999	Fernwaerme	0.246

Jahr	Energietraeger	Emissionen
1999	Erdgas	1.319
1999	Heizoel_EL	1.728
2000	Strom	0.790
2000	Kerosin	0.833
2000	Diesel	0.432
2000	Benzin	1.095
2000	Holz_UW_BG_SK	0.021
2000	Fernwaerme	0.248
2000	Erdgas	1.353
2000	Heizoel_EL	1.650
2001	Strom	1.091
2001	Kerosin	0.782
2001	Diesel	0.445
2001	Benzin	1.080
2001	Holz_UW_BG_SK	0.019
2001	Fernwaerme	0.250

Jahr	Energietraeger	Emissionen
2001	Erdgas	1.450
2001	Heizoel_EL	1.189
2002	Strom	1.381
2002	Kerosin	0.719
2002	Diesel	0.453
2002	Benzin	1.057
2002	Holz_UW_BG_SK	0.019
2002	Fernwaerme	0.251
2002	Erdgas	1.374
2002	Heizoel_EL	1.498
2003	Strom	1.169
2003	Kerosin	0.642
2003	Diesel	0.470
2003	Benzin	1.032
2003	Holz_UW_BG_SK	0.018
2003	Fernwaerme	0.286

Jahr	Energietraeger	Emissionen
2003	Erdgas	1.490
2003	Heizoel_EL	1.317
2004	Strom	0.954
2004	Kerosin	0.601
2004	Diesel	0.486
2004	Benzin	1.003
2004	Holz_UW_BG_SK	0.019
2004	Fernwaerme	0.298
2004	Erdgas	1.508
2004	Heizoel_EL	1.305
2005	Strom	0.170
2005	Kerosin	0.605
2005	Diesel	0.503
2005	Benzin	0.970
2005	Holz_UW_BG_SK	0.020
2005	Fernwaerme	0.300

Jahr	Energietraeger	Emissionen
2005	Erdgas	1.602
2005	Heizoel_EL	1.281
2006	Strom	0.176
2006	Kerosin	0.629
2006	Diesel	0.504
2006	Benzin	0.889
2006	Holz_UW_BG_SK	0.019
2006	Fernwaerme	0.255
2006	Erdgas	1.450
2006	Heizoel_EL	1.425
2007	Strom	0.176
2007	Kerosin	0.665
2007	Diesel	0.508
2007	Benzin	0.809
2007	Holz_UW_BG_SK	0.017
2007	Fernwaerme	0.218

Jahr	Energietraeger	Emissionen
2007	Erdgas	1.347
2007	Heizoel_EL	1.210
2008	Strom	0.180
2008	Kerosin	0.701
2008	Diesel	0.507
2008	Benzin	0.738
2008	Holz_UW_BG_SK	0.018
2008	Fernwaerme	0.243
2008	Erdgas	1.409
2008	Heizoel_EL	1.155
2009	Strom	0.183
2009	Kerosin	0.666
2009	Diesel	0.516
2009	Benzin	0.683
2009	Holz_UW_BG_SK	0.020
2009	Fernwaerme	0.227

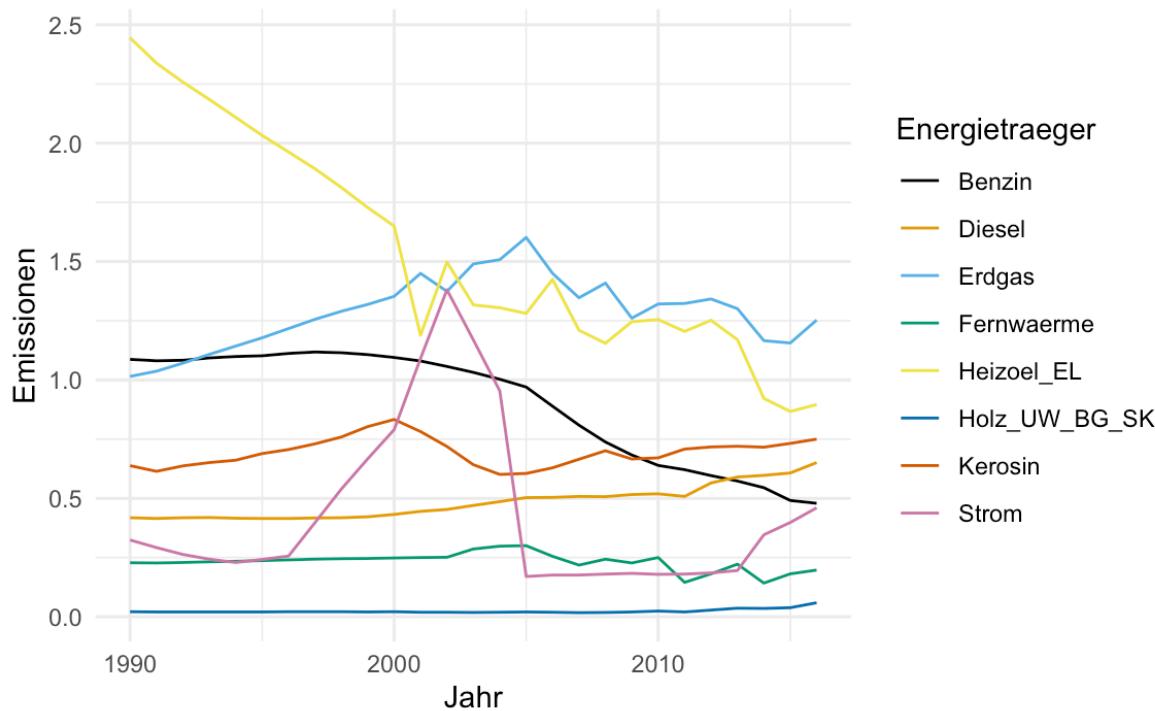
Jahr	Energietraeger	Emissionen
2009	Erdgas	1.261
2009	Heizoel_EL	1.246
2010	Strom	0.179
2010	Kerosin	0.671
2010	Diesel	0.519
2010	Benzin	0.639
2010	Holz_UW_BG_SK	0.024
2010	Fernwaerme	0.250
2010	Erdgas	1.321
2010	Heizoel_EL	1.255
2011	Strom	0.180
2011	Kerosin	0.708
2011	Diesel	0.508
2011	Benzin	0.621
2011	Holz_UW_BG_SK	0.020
2011	Fernwaerme	0.145

Jahr	Energietraeger	Emissionen
2011	Erdgas	1.323
2011	Heizoel_EL	1.205
2012	Strom	0.185
2012	Kerosin	0.717
2012	Diesel	0.565
2012	Benzin	0.596
2012	Holz_UW_BG_SK	0.028
2012	Fernwaerme	0.181
2012	Erdgas	1.342
2012	Heizoel_EL	1.252
2013	Strom	0.195
2013	Kerosin	0.720
2013	Diesel	0.590
2013	Benzin	0.573
2013	Holz_UW_BG_SK	0.036
2013	Fernwaerme	0.222

Jahr	Energietraeger	Emissionen
2013	Erdgas	1.301
2013	Heizoel_EL	1.170
2014	Strom	0.346
2014	Kerosin	0.716
2014	Diesel	0.597
2014	Benzin	0.545
2014	Holz_UW_BG_SK	0.035
2014	Fernwaerme	0.142
2014	Erdgas	1.166
2014	Heizoel_EL	0.922
2015	Strom	0.398
2015	Kerosin	0.732
2015	Diesel	0.607
2015	Benzin	0.491
2015	Holz_UW_BG_SK	0.038
2015	Fernwaerme	0.181

Drei Variablen -> Drei visuelle Eigenschaften (aesthetics)

```
1 ggplot(data = ghg_long,
2         mapping = aes(x = Jahr,
3                         y = Emissionen,
4                         color = Energietraeger)) +
5   geom_line() +
6   scale_color_colorblind() +
7   theme_minimal()
```



Wie wird's gemacht?

1 ghg

Jahr	Strom	Kerosin	Diesel	Benzin	Holz_UW_BG_SK	Fernwaerme	Erdgas	Heizoel
1990	0.324	0.638	0.418	1.087	0.021	0.228	1.015	2.4
1991	0.292	0.614	0.415	1.081	0.020	0.227	1.037	2.3
1992	0.263	0.637	0.418	1.083	0.020	0.229	1.071	2.1
1993	0.243	0.651	0.419	1.093	0.020	0.232	1.108	2.0
1994	0.229	0.661	0.416	1.099	0.020	0.234	1.143	2.0
1995	0.242	0.689	0.415	1.102	0.020	0.237	1.178	2.0
1996	0.256	0.706	0.415	1.112	0.021	0.240	1.217	1.9
1997	0.397	0.730	0.417	1.118	0.021	0.243	1.256	1.8
1998	0.539	0.759	0.418	1.115	0.021	0.245	1.290	1.7
1999	0.667	0.803	0.422	1.107	0.020	0.246	1.319	1.6
2000	0.790	0.833	0.432	1.095	0.021	0.248	1.353	1.6
2001	1.091	0.782	0.445	1.080	0.019	0.250	1.450	1.5
2002	1.381	0.719	0.453	1.057	0.019	0.251	1.374	1.4

Jahr	Strom	Kerosin	Diesel	Benzin	Holz_UW_BG_SK	Fernwaerme	Erdgas	Heizoel
2003	1.169	0.642	0.470	1.032		0.018	0.286	1.490
2004	0.954	0.601	0.486	1.003		0.019	0.298	1.508
2005	0.170	0.605	0.503	0.970		0.020	0.300	1.602
2006	0.176	0.629	0.504	0.889		0.019	0.255	1.450
2007	0.176	0.665	0.508	0.809		0.017	0.218	1.347
2008	0.180	0.701	0.507	0.738		0.018	0.243	1.409
2009	0.183	0.666	0.516	0.683		0.020	0.227	1.261
2010	0.179	0.671	0.519	0.639		0.024	0.250	1.321
2011	0.180	0.708	0.508	0.621		0.020	0.145	1.323
2012	0.185	0.717	0.565	0.596		0.028	0.181	1.342
2013	0.195	0.720	0.590	0.573		0.036	0.222	1.301
2014	0.346	0.716	0.597	0.545		0.035	0.142	1.166
2015	0.398	0.732	0.607	0.491		0.038	0.181	1.156
2016	0.460	0.750	0.651	0.479		0.059	0.197	1.253

Wie wird's gemacht?

```

1 ghg |>
2   pivot_longer(cols = !Jahr,
3                 names_to = "Energietraeger",
4                 values_to = "Emissionen")

```

Jahr	Energietraeger	Emissionen
1990	Strom	0.324
1990	Kerosin	0.638
1990	Diesel	0.418
1990	Benzin	1.087
1990	Holz_UW_BG_SK	0.021
1990	Fernwaerme	0.228
1990	Erdgas	1.015
1990	Heizoel_EL	2.445
1991	Strom	0.292
1991	Kerosin	0.614
1991	Diesel	0.415
1991	Benzin	1.081
1991	Holz_UW_BG_SK	0.020
1991	Fernwaerme	0.227
1991	Erdgas	1.037
1991	Heizoel_EL	2.338
1992	Strom	0.263

1992	Strom	0.268
1992	Kerosin	0.637
1992	Diesel	0.418
1992	Benzin	1.083
1992	Holz_UW_BG_SK	0.020
1992	Fernwaerme	0.229
1992	Erdgas	1.071
1992	Heizoel_EL	2.258
1993	Strom	0.243
1993	Kerosin	0.651
1993	Diesel	0.419
1993	Benzin	1.093
1993	Holz_UW_BG_SK	0.020
1993	Fernwaerme	0.232
1993	Erdgas	1.108
1993	Heizoel_EL	2.185
1994	Strom	0.229
1994	Kerosin	0.661

1994	Diesel	0.416
1994	Benzin	1.099
1994	Holz_UW_BG_SK	0.020
1994	Fernwaerme	0.234
1994	Erdgas	1.143
1994	Heizoel_EL	2.109
1995	Strom	0.242
1995	Kerosin	0.689
1995	Diesel	0.415
1995	Benzin	1.102
1995	Holz_UW_BG_SK	0.020
1995	Fernwaerme	0.237
1995	Erdgas	1.178
1995	Heizoel_EL	2.033
1996	Strom	0.256
1996	Kerosin	0.706
1996	Diesel	0.415
1996	Benzin	1.112

1996	Holz_UW_BG_SK	0.021
1996	Fernwaerme	0.240
1996	Erdgas	1.217
1996	Heizoel_EL	1.963
1997	Strom	0.397
1997	Kerosin	0.730
1997	Diesel	0.417
1997	Benzin	1.118
1997	Holz_UW_BG_SK	0.021
1997	Fernwaerme	0.243
1997	Erdgas	1.256
1997	Heizoel_EL	1.892
1998	Strom	0.539
1998	Kerosin	0.759
1998	Diesel	0.418
1998	Benzin	1.115
1998	Holz_UW_BG_SK	0.021

1998	Fernwaerme	0.245
1998	Erdgas	1.290
1998	Heizoel_EL	1.813
1999	Strom	0.667
1999	Kerosin	0.803
1999	Diesel	0.422
1999	Benzin	1.107
1999	Holz_UW_BG_SK	0.020
1999	Fernwaerme	0.246
1999	Erdgas	1.319
1999	Heizoel_EL	1.728
2000	Strom	0.790
2000	Kerosin	0.833
2000	Diesel	0.432
2000	Benzin	1.095
2000	Holz_UW_BG_SK	0.021
2000	Fernwaerme	0.248
2000	Erdgas	1.353

2000	Heizoel_EL	1.650
2001	Strom	1.091
2001	Kerosin	0.782
2001	Diesel	0.445
2001	Benzin	1.080
2001	Holz_UW_BG_SK	0.019
2001	Fernwaerme	0.250
2001	Erdgas	1.450
2001	Heizoel_EL	1.189
2002	Strom	1.381
2002	Kerosin	0.719
2002	Diesel	0.453
2002	Benzin	1.057
2002	Holz_UW_BG_SK	0.019
2002	Fernwaerme	0.251
2002	Erdgas	1.374
2002	Heizoel_EL	1.498
2003	Strom	1.160

2003	Strom	1.169
2003	Kerosin	0.642
2003	Diesel	0.470
2003	Benzin	1.032
2003	Holz_UW_BG_SK	0.018
2003	Fernwaerme	0.286
2003	Erdgas	1.490
2003	Heizoel_EL	1.317
2004	Strom	0.954
2004	Kerosin	0.601
2004	Diesel	0.486
2004	Benzin	1.003
2004	Holz_UW_BG_SK	0.019
2004	Fernwaerme	0.298
2004	Erdgas	1.508
2004	Heizoel_EL	1.305
2005	Strom	0.170
2005	Kerosin	0.605

2005	Diesel	0.503
2005	Benzin	0.970
2005	Holz_UW_BG_SK	0.020
2005	Fernwaerme	0.300
2005	Erdgas	1.602
2005	Heizoel_EL	1.281
2006	Strom	0.176
2006	Kerosin	0.629
2006	Diesel	0.504
2006	Benzin	0.889
2006	Holz_UW_BG_SK	0.019
2006	Fernwaerme	0.255
2006	Erdgas	1.450
2006	Heizoel_EL	1.425
2007	Strom	0.176
2007	Kerosin	0.665
2007	Diesel	0.508
2007	Benzin	0.809

2007	Benzin	0.337
2007	Holz_UW_BG_SK	0.017
2007	Fernwaerme	0.218
2007	Erdgas	1.347
2007	Heizoel_EL	1.210
2008	Strom	0.180
2008	Kerosin	0.701
2008	Diesel	0.507
2008	Benzin	0.738
2008	Holz_UW_BG_SK	0.018
2008	Fernwaerme	0.243
2008	Erdgas	1.409
2008	Heizoel_EL	1.155
2009	Strom	0.183
2009	Kerosin	0.666
2009	Diesel	0.516
2009	Benzin	0.683
2009	Holz_UW_BG_SK	0.020

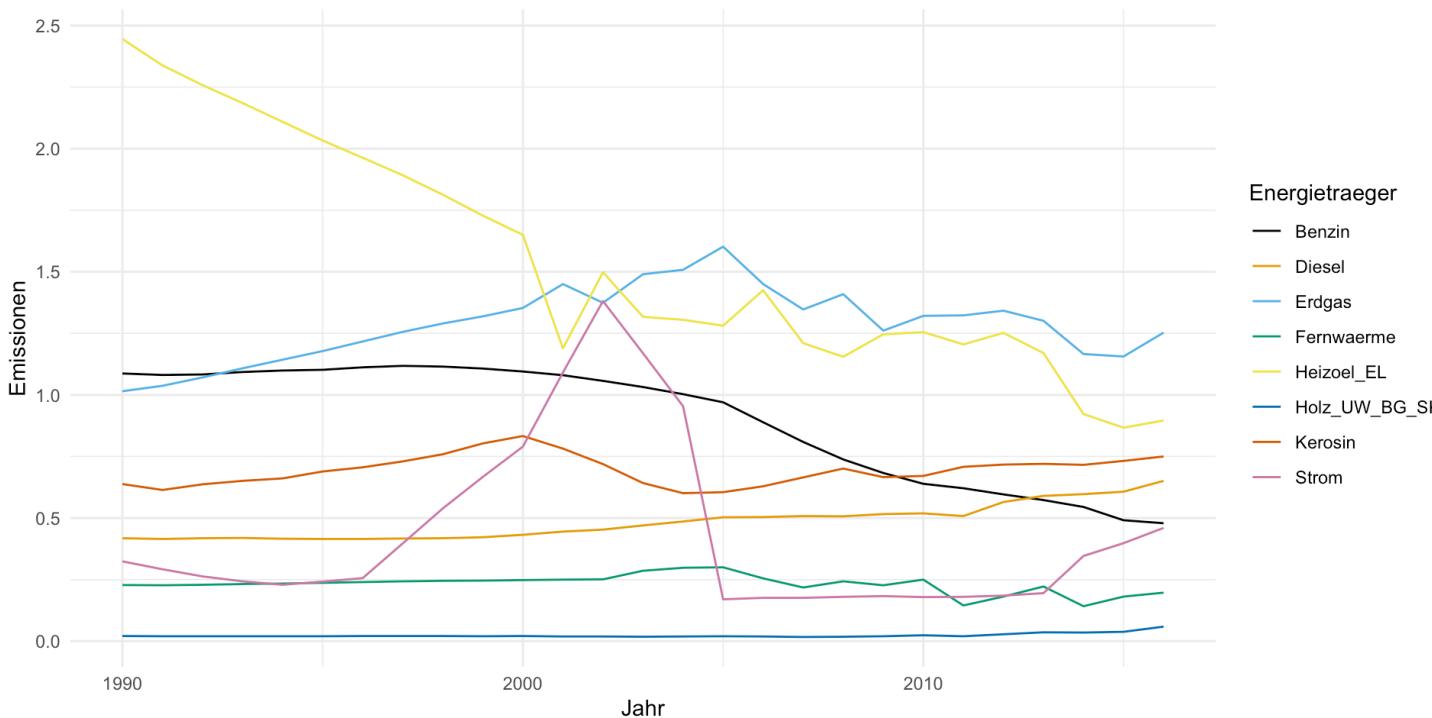
2009	Fernwaerme	0.227
2009	Erdgas	1.261
2009	Heizoel_EL	1.246
2010	Strom	0.179
2010	Kerosin	0.671
2010	Diesel	0.519
2010	Benzin	0.639
2010	Holz_UW_BG_SK	0.024
2010	Fernwaerme	0.250
2010	Erdgas	1.321
2010	Heizoel_EL	1.255
2011	Strom	0.180
2011	Kerosin	0.708
2011	Diesel	0.508
2011	Benzin	0.621
2011	Holz_UW_BG_SK	0.020
2011	Fernwaerme	0.145
2011	Erdgas	1.323

2011	Heizoel_EL	1.205
2012	Strom	0.185
2012	Kerosin	0.717
2012	Diesel	0.565
2012	Benzin	0.596
2012	Holz_UW_BG_SK	0.028
2012	Fernwaerme	0.181
2012	Erdgas	1.342
2012	Heizoel_EL	1.252
2013	Strom	0.195
2013	Kerosin	0.720
2013	Diesel	0.590
2013	Benzin	0.573
2013	Holz_UW_BG_SK	0.036
2013	Fernwaerme	0.222
2013	Erdgas	1.301
2013	Heizoel_EL	1.170
2014	Strom	0.216

2014	Strom	0.346
2014	Kerosin	0.716
2014	Diesel	0.597
2014	Benzin	0.545
2014	Holz_UW_BG_SK	0.035
2014	Fernwaerme	0.142
2014	Erdgas	1.166
2014	Heizoel_EL	0.922
2015	Strom	0.398
2015	Kerosin	0.732
2015	Diesel	0.607
2015	Benzin	0.491
2015	Holz_UW_BG_SK	0.038
2015	Fernwaerme	0.181
2015	Erdgas	1.156
2015	Heizoel_EL	0.867
2016	Strom	0.460
2016	Kerosin	0.750

Drei Variablen -> Drei visuelle Eigenschaften (aesthetics)

```
1 ggplot(data = ghg_long,
2         mapping = aes(x = Jahr,
3                         y = Emissionen,
4                         color = Energietraeger)) +
5   geom_line() +
6   scale_color_colorblind() +
7   theme_minimal()
```



Tidy data

Eigenschaften von Tidy data

- Eigenschaft 1: Jede Spalte ist eine Variable
- Eigenschaft 2: Jede Reihe ist eine Beobachtung
- Eigenschaft 3: Jede Zelle enthält eine Messung

Penguins

Erfüllen die Daten die Eingeschaften für Tidy data?

species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year
Adelie	Torgersen	39.1	18.7	181	3750	male	2007
Adelie	Torgersen	39.5	17.4	186	3800	female	2007
Adelie	Torgersen	40.3	18.0	195	3250	female	2007
Adelie	Torgersen	NA	NA	NA	NA	NA	2007
Adelie	Torgersen	36.7	19.3	193	3450	female	2007
Adelie	Torgersen	39.3	20.6	190	3650	male	2007
Adelie	Torgersen	38.9	17.8	181	3625	female	2007
Adelie	Torgersen	39.2	19.6	195	4675	male	2007
Adelie	Torgersen	34.1	18.1	193	3475	NA	2007
Adelie	Torgersen	42.0	20.2	190	4250	NA	2007

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

	A	B	C	D	E	F	G	H	I
1	cc-d-01.02.04.02		Ständige Wohnbevölkerung nach Altersklasse und Altersmasszahlen nach Kanton, am 31.12.2020						
2			Provisorische Jahresergebnisse						
3	Grossregionen Kantone	Total	0-19 Jahre	20-39 Jahre	40-64 Jahre	65-79 Jahre	80 Jahre und mehr	Jugendquotient ¹	Altersquotient ²
4	Schweiz	8 667 088	1 723 565	2 280 707	3 032 787	1 171 506	458 523	32.4	30.7
5	Genferseeregion	1 668 471	351 278	455 418	574 194	205 266	82 315	34.1	27.9
6	Waadt	814 075	177 447	225 069	276 942	96 173	38 444	35.3	26.8
7	Wallis	348 318	67 628	89 388	121 294	51 839	18 169	32.1	33.2
8	Genf	506 078	106 203	140 961	175 958	57 254	25 702	33.5	26.2
9	Espace Mittelland	1 894 879	373 753	481 384	659 098	275 838	104 806	32.8	33.4
10	Bern	1 042 516	197 760	260 221	362 072	160 402	62 061	31.8	35.7
11	Freiburg	325 419	71 505	88 526	112 785	39 646	12 957	35.5	26.1
12	Solothurn	277 396	52 800	69 973	98 981	40 445	15 197	31.3	32.9
13	Neuenburg	175 860	36 459	44 834	60 437	23 984	10 146	34.6	32.4
14	Jura	73 688	15 229	17 830	24 823	11 361	4 445	35.7	37.1
15	Nordwestschweiz	1 181 397	230 093	302 831	418 419	165 110	64 944	31.9	31.9

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

	A	B	C	D	E	F	G	H	I
1	cc-d-01.02.04.02	Ständige Wohnbevölkerung	Variable Altersgruppe als Reihe						
2		Provisorische Jahresergebnisse							
3	Grossregionen Kantone	Total	0-19 Jahre	20-39 Jahre	40-64 Jahre	65-79 Jahre	80 Jahre und mehr	Jugendquotient ¹	Altersquotient ²
4	Schweiz	8 667 088	1 723 565	2 280 707	3 032 787	1 171 506	458 523	32.4	30.7
5	Genferseeregion	1 668 471	351 278	455 418	574 194	205 266	82 315	34.1	27.9
6	Waadt	814 075	177 447	225 069	276 942	96 173	38 444	35.3	26.8
7	Wallis	348 318	67 628	89 388	121 294	51 839	18 169	32.1	33.2
8	Genf	506 078	106 203	140 961	175 958	57 254	25 702	33.5	26.2
9	Espace Mittelland	1 894 879	373 753	481 384	659 098	275 838	104 806	32.8	33.4
10	Bern	1 042 516	197 760	260 221	362 072	160 402	62 061	31.8	35.7
11	Freiburg	325 419	71 505	88 526	112 785	39 646	12 957	35.5	26.1
12	Solothurn	277 396	52 800	69 973	98 981	40 445	15 197	31.3	32.9
13	Neuenburg	175 860	36 459	44 834	60 437	23 984	10 146	34.6	32.4
14	Jura	73 688	15 229	17 830	24 823	11 361	4 445	35.7	37.1
15	Nordwestschweiz	1 181 397	230 093	302 831	418 419	165 110	64 944	31.9	31.9

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

	A	B	C	D	E	F	G	H	I
1	cc-d-01.02.04.02	Ständige Wohnbevölkerung	Variable Altersgruppe als Reihe						
2		Provisorische Jahresergebnisse							
3	Grossregionen Kantone	Total	0-19 Jahre	20-39 Jahre	40-64 Jahre	65-79 Jahre	80 Jahre und mehr	Jugendquotient ¹	Altersquotient ²
4	Schweiz	8 667 088	1 723 565	2 280 707	3 032 787	1 171 506	458 523	32.4	30.7
5	Genferseeregion	1 668 471	351 278	455 418	574 194	205 266	82 315	34.1	27.9
6	Waadt	814 075	177 447	225 069	276 942	96 173	38 444	35.3	26.8
7	Wallis	348 318	67 628	89 388	121 294	51 839	18 169	32.1	33.2
8	Genf	506 078	106 203	140 961	175 958	57 254	25 702	33.5	26.2
9	Espace Mittelland	1 894 879	373 753	481 384	659 098	275 838	104 806	32.8	33.4
10	Bern	1 042 516	197 760	260 221	362 072	160 402	62 061	31.8	35.7
11	Freiburg	325 419	71 505	88 526	112 785	39 646	12 957	35.5	26.1
12	Solothurn	277 396	52 800	69 973	98 981	40 445	15 197	31.3	32.9
13	Neuenburg	175 860	36 459	44 834	60 437	23 984	10 146	34.6	32.4
14	Jura	73 688	15 229	17 830	24 823	11 361	4 445	35.7	37.1
15	Nordwestschweiz	1 181 397	230 093	302 831	418 419	165 110	64 944	31.9	31.9

Reihen als Zusammenfassung (Summe)

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

	A	B	C	H	M	R	S	T	U	V	W	X	Y	Z	
1	Kosten des Gesundheitswesens nach Leistungen 1)													T 14.5.1.3	
2	In Millionen Franken														
3															
4															
5	Total		1995	2000	2005	2010	2011	2012	2013	2014	2015	2016	2017	2018	
6	L Stationäre Kurativbehandlung 2)		36 056	43 072	52 388	62 565	64 243	66 512	69 118	71 429	74 385	77 455	79 643	80 242	
7	L1 Stationäre somatische Akutbehandlung 2)		9 742	10 786	12 584	13 373	13 583	14 176	14 791	14 947	15 386	15 758	15 718	15 548	
8	L2 Stationäre Psychiatriebehandlung 2)						11 696	11 878	12 397	12 946	13 118	13 469	13 832	13 786	13 622
9	L3 Stationäre Geburtshausbehandlung 2)						1 674	1 699	1 771	1 836	1 819	1 905	1 912	1 917	1 908
10	M Ambulante Kurativbehandlung						3	5	8	9	10	12	14	16	17
11	M2 Ambulante somatische Akutbehandlung im Spital		8 336	10 243	12 699	15 808	16 109	16 924	17 688	18 681	19 541	20 436	21 108	20 753	
12	M3 Ärztliche Behandlung, ambulant, Einzelleistungen 3)						4 226	4 315	4 717	4 969	5 427	5 677	6 136	6 307	6 409
13	M4 Ärztliche Behandlung, ambulant, Managed Care 3)						4 509	4 317	4 273	4 343	4 405	4 638	4 711	4 690	3 871
14	M6 Zahnbehandlung						1 582	1 984	2 277	2 578	2 839	3 195	3 400	3 659	3 797
15	M7 Ambulante Psychiatrie- und Psychologiebehandlung, kurativ 4)						4 022	4 089	4 171	4 251	4 347	4 279	4 256	4 473	4 684
							864	854	923	987	1 121	1 169	1 263	1 301	1 391

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

	A	B	C	H	M	R	S	T	U	V	W	X	Y	Z
1	Kosten des Gesundheitswesens nach Leistungen 1)													T 14.5.1.3
2	In Millionen Franken													
3														
4														
5	Total		1995	2000	2005	2010	2011	2012	2013	2014	2015	2016	2017	2018
6	L Stationäre Kurativbehandlung 2)		36 056	43 072	52 388	62 565	64 243	66 512	69 118	71 429	74 385	77 455	79 643	80 242
7	L1 Stationäre somatische Akutbehandlung 2)		9 742	10 786	12 584	13 373	13 583	14 176	14 791	14 947	15 386	15 758	15 718	15 548
8	L2 Stationäre Psychiatriebehandlung 2)					11 696	11 878	12 397	12 946	13 118	13 469	13 832	13 786	13 622
9	L3 Stationäre Geburtshausbehandlung 2)					1 674	1 699	1 771	1 836	1 819	1 905	1 912	1 917	1 908
10	M Ambulante Kurativbehandlung					3	5	8	9	10	12	14	16	17
11	M2 Ambulante somatische Akutbehandlung im Spital		8 336	10 243	12 699	15 808	16 109	16 924	17 688	18 681	19 541	20 436	21 108	20 753
12	M3 Ärztliche Behandlung, ambulant, Einzelleistungen 3)					4 226	4 315	4 717	4 969	5 427	5 677	6 136	6 307	6 409
13	M4 Ärztliche Behandlung, ambulant, Managed Care 3)					4 509	4 317	4 273	4 343	4 405	4 638	4 711	4 690	3 871
14	M6 Zahnbehandlung					1 582	1 984	2 277	2 578	2 839	3 195	3 400	3 659	3 797
15	M7 Ambulante Psychiatrie- und Psychologiebehandlung, kurativ 4)					4 022	4 089	4 171	4 251	4 347	4 279	4 256	4 473	4 684
						864	854	923	987	1 121	1 169	1 263	1 301	1 391

Variable Jahr als Reihe

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

	A	B	C	H	M	R	S	T	U	V	W	X	Y	Z
1	Kosten des Gesundheitswesens nach Leistungen 1)													
2	In Millionen Franken													
3														
4														
5	Total		1995	2000	2005	2010	2011	2012	2013	2014	2015	2016	2017	2018
6	L Stationäre Kurativbehandlung 2)		36 056	43 072	52 388	62 565	64 243	66 512	69 118	71 429	74 385	77 455	79 643	80 242
7	L1 Stationäre somatische Akutbehandlung 2)		9 742	10 786	12 584	13 373	13 583	14 176	14 791	14 947	15 386	15 758	15 718	15 548
8	L2 Stationäre Psychiatriebehandlung 2)					11 696	11 676	12 597	12 946	13 116	13 469	13 652	13 766	13 622
9	L3 Stationäre Geburtshausbehandlung 2)					1 674	1 699	1 771	1 836	1 819	1 905	1 912	1 917	1 908
10	M Ambulante Kurativbehandlung					3	5	8	9	10	12	14	16	17
11	M2 Ambulante somatische Akutbehandlung im Spital		8 336	10 243	12 699	15 808	16 109	16 924	17 688	18 681	19 541	20 436	21 108	20 753
12	M3 Ärztliche Behandlung, ambulant, Einzelleistungen 3)					4 226	4 315	4 717	4 969	5 427	5 677	6 136	6 307	6 409
13	M4 Ärztliche Behandlung, ambulant, Managed Care 3)					4 509	4 317	4 273	4 343	4 405	4 638	4 711	4 690	3 871
14	M6 Zahnbehandlung					1 582	1 984	2 277	2 578	2 839	3 195	3 400	3 659	3 797
15	M7 Ambulante Psychiatrie- und Psychologiebehandlung, kurativ 4)					4 022	4 089	4 171	4 251	4 347	4 279	4 256	4 473	4 684
						864	854	923	987	1 121	1 169	1 263	1 301	1 391

Variable Jahr als Reihe

Reihen als Zusammenfassung (Summe)

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

	A	B	C	D	E	F	G	H
1	Klimadaten: Eistage							
4	Jahr	Anzahl Eistage (maximale Temperatur < 0 °C) und langjähriger Mittelwert 1961-1990 (Ø)						
6	Basel-Binningen		Bern-Zollikofen 2)	Davos	Genf-Cointrin	Locarno-Monti	Lugano	Luzern
7	316 m ü. M.		553 m ü. M.	1594 m ü. M.	411 m ü. M.	367 m ü. M.	273 m ü. M.	454 m ü. M.
8	Ø 15,9 Tage		Ø 24,9 Tage	Ø 64,1 Tage	Ø 11,9 Tage	Ø 1,5 Tag	Ø 0,7 Tag	Ø 21,3 Tage
11	1959	7	17	39	3	0	0	11
12	1960	18	23	42	11	2	2	17
13	1961	14	18	44	5	3	0	14
14	1962	22	29	86	15	2	2	25
15	1963	48	54	78	45	11	6	52
16	1964	21	38	53	18	0	0	30
17	1965	12	17	85	5	0	0	15
18	1966	15	17	78	11	1	0	17
		Eistage	Frosttage	Sommertage	Hitzetage	Tropennächte	Niederschlagstage	+

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

	A	B	C	D	E	F	G	H
1	Klimadaten: Eistage							
4	Jahr	Anzahl Eistage (maximale Temperatur < 0 °C) und langjähriger Mittelwert 1961-1990 (Ø)						
6	Basel-Binningen		Bern-Zollikofen 2)	Davos	Genf-Cointrin	Locarno-Monti	Lugano	Luzern
7	316 m ü. M.		553 m ü. M.	1594 m ü. M.	411 m ü. M.	367 m ü. M.	273 m ü. M.	454 m ü. M.
8	Ø 15,9 Tage		Ø 24,9 Tage	Ø 64,1 Tage	Ø 11,9 Tage	Ø 1,5 Tag	Ø 0,7 Tag	Ø 21,3 Tage
11	1959	7	17	39	3	0	0	11
12	1960	18	23	42	11	2	2	17
13	1961	14	18	44	5	3	0	14
14	1962	22	29	86	15	2	2	25
15	1963	48	54	78	45	11	6	52
16	1964	21	38	53	18	0	0	30
17	1965	12	17	85	5	0	0	15
18	1966	15	17	78	11	1	0	17
Eistage Frosttage Sommertage Hitzetage Tropennächte Niederschlagstage + - 160%								

Variable als Worksheets

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

Klimadaten: Eistage

Definitionen in Reihe

A	B	C	D	E	F	G	H
1	Klimadaten: Eistage						
4	Jahr	Anzahl Eistage (maximale Temperatur < 0 °C) und langjähriger Mittelwert 1961-1990 (Ø)	1				
6	Basel-Binningen	Bern-Zollikofen 2)	Davos	Genf-Cointrin	Locarno-Monti	Lugano	Luzern
7	316 m ü. M.	553 m ü. M.	1594 m ü. M.	411 m ü. M.	367 m ü. M.	273 m ü. M.	454 m ü. M.
8	Ø 15,9 Tage	Ø 24,9 Tage	Ø 64,1 Tage	Ø 11,9 Tage	Ø 1,5 Tag	Ø 0,7 Tag	Ø 21,3 Tage
11	1959	7	17	39	3	0	0
12	1960	18	23	42	11	2	2
13	1961	14	18	44	5	3	0
14	1962	22	29	86	15	2	2
15	1963	48	54	78	45	11	6
16	1964	21	38	53	18	0	0
17	1965	12	17	85	5	0	0
18	1966	15	17	78	11	1	0
		Eistage	Frosttage	Sommertage	Hitzetage	Tropennächte	Niederschlagstage

Variable als Worksheets

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

A Klimadaten: Eistage C Definitionen in Reihe F Variablen als Reihen

1	B	C	D	E	F	G	H
4	Jahr	Anzahl Eistage (maximale Temperatur < 0 °C) und langjähriger Mittelwert 1961-1990 (Ø) 1)					
6	Basel-Binningen	Bern-Zollikofen 2)	Davos	Genf-Cointrin	Locarno-Monti	Lugano	Luzern
7	316 m ü. M.	553 m ü. M.	1594 m ü. M.	411 m ü. M.	367 m ü. M.	273 m ü. M.	454 m ü. M.
8	Ø 15,9 Tage	Ø 24,9 Tage	Ø 64,1 Tage	Ø 11,9 Tage	Ø 1,5 Tag	Ø 0,7 Tag	Ø 21,3 Tage
11	1959	7	17	39	3	0	0
12	1960	18	23	42	11	2	2
13	1961	14	18	44	5	3	0
14	1962	22	29	86	15	2	2
15	1963	48	54	78	45	11	6
16	1964	21	38	53	18	0	0
17	1965	12	17	85	5	0	0
18	1966	15	17	78	11	1	0
Eistage Frosttage Sommertage Hitzetage Tropennächte Niederschlagstage +							
160%							
Variable als Worksheets							

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

Klimadaten: Eistage

Definitionen in Reihe

Variablen als Reihen

Reihe als Zusammenfassung (Durchschnitt)

Eistage

A	B	C	D	E	F	G	H
1	Klimadaten: Eistage						
4	Jahr	Anzahl Eistage (maximale Temperatur < 0 °C) und langjähriger Mittelwert 1961-1990 (Ø) 1					
6	Basel-Binningen	Bern-Zollikofen 2)	Davos	Genf-Cointrin	Locarno-Monti	Lugano	Luzern
7	316 m ü. M.	553 m ü. M.	1594 m ü. M.	411 m ü. M.	367 m ü. M.	273 m ü. M.	454 m ü. M.
8	Ø 15,9 Tage	Ø 24,9 Tage	Ø 64,1 Tage	Ø 11,9 Tage	Ø 1,5 Tag	Ø 0,7 Tag	Ø 21,3 Tage
11	1959	7	17	20	2	0	0
12	1960	18	17	20	2	0	11
13	1961	14	18	44	5	3	0
14	1962	22	29	86	15	2	2
15	1963	48	54	78	45	11	6
16	1964	21	38	53	18	0	0
17	1965	12	17	85	5	0	0
18	1966	15	17	78	11	1	0
							17
	Eistage	Frosttage	Sommertage	Hitzetage	Tropennächte	Niederschlagstage	+

Variable als Worksheets

- Welche Eigenschaften von Tidy data sind hier nicht erfüllt?

???

	A	B	C	D	E	F	G	H
1	Klimadaten: Eistage	Definitionen in Reihe					Variablen als Reihen	
4	Jahr	Anzahl Eistage (maximale Temperatur < 0 °C) und langjähriger Mittelwert 1961-1990 (Ø)						
6		Basel-Binningen	Bern-Zollikofen 2)	Davos	Genf-Cointrin	Locarno-Monti	Lugano	Luzern
7		316 m ü. M.	553 m ü. M.	1594 m ü. M.	411 m ü. M.	367 m ü. M.	273 m ü. M.	454 m ü. M.
8		Ø 15,9 Tage	Ø 24,9 Tage	Ø 64,1 Tage	Ø 11,9 Tage	Ø 1,5 Tag	Ø 0,7 Tag	Ø 21,3 Tage
11	1959	7	17	20	2	0	0	11
12	1960	18	Reihe als Zusammenfassung (Durchschnitt)					
13	1961	14	18	44	5	3	0	14
14	1962	22	29	86	15	2	2	25
15	1963	48	54	78	45	11	6	52
16	1964	21	38	53	18	0	0	30
17	1965	12	17	85	5	0	0	15
18	1966	15	17	78	11	1	0	17
<input type="button" value="Eistage"/> <input type="button" value="Frosttage"/> <input type="button" value="Sommertage"/> <input type="button" value="Hitzetage"/> <input type="button" value="Tropennächte"/> <input type="button" value="Niederschlagstage"/> <input type="button" value="+"/>								
Variable als Worksheets								

Relevanter Unterschied - Ziel der Daten Publikation

Daten in Tabellen darstellen

- Layout
 - Gut leserlich
 - Kompakt
 - Erkenntnis bringend
- Metadaten

Daten für weitere Nutzung bereitstellen

- Layout (Tidy data)
 - Eigenschaft 1: Jede Spalte ist eine Variable
 - Eigenschaft 2: Jede Reihe ist eine Beobachtung
 - Eigenschaft 3: Jede Zelle enthält eine Messung
- Keine Metadaten
- Keine Farben, Formatierungen, etc.
- Folgt Standards (Datum: ISO 8601)
- etc.

Pause machen

Bitte steh auf und beweg dich. Lasst eure E-Mails in Frieden ruhen.

Wir sind dran: 01-treibhausgase-ihr.qmd

1. Öffne [posit.cloud](#) in deinem Browser (verwende dein Lesezeichen).
2. Öffne den rstatszh-k010 Arbeitsbereich (Workspace) für den Kurs.
3. Klicke auf **Continue** neben **md-07-uebungen**.
4. Suche im Dateimanager im Fenster unten rechts die Datei **01-treibhausgase-ihr.qmd** und klicke darauf, um sie im Fenster oben links zu öffnen.

Teil 2: Daten zusammenfügen

Mit mehreren Dataframes arbeiten

Wir haben mehrere Dataframes und wollen diese zusammenbringen:

```
1 professions <- read_csv("daten/scientists/professions.csv")  
2 dates <- read_csv("daten/scientists/dates.csv")  
3 works <- read_csv("daten/scientists/works.csv")
```

Daten: Women in science

Informationen zu 10 Frauen in der Wissenschaft welche die Welt verändert haben

name

Ada Lovelace

Marie Curie

Janaki Ammal

Chien-Shiung Wu

Katherine Johnson

Rosalind Franklin

Vera Rubin

name

Gladys West

Flossie Wong-Staal

Jennifer Doudna

Inputs

professions dates works

```
# A tibble: 10 × 2
  name      profession
  <chr>     <chr>
1 Ada Lovelace Mathematician
2 Marie Curie Physicist and Chemist
3 Janaki Ammal Botanist
4 Chien-Shiung Wu Physicist
5 Katherine Johnson Mathematician
6 Rosalind Franklin Chemist
7 Vera Rubin Astronomer
8 Gladys West Mathematician
9 Flossie Wong-Staal Virologist and Molecular Biologist
10 Jennifer Doudna Biochemist
```

Gewünschter Output

```
# A tibble: 10 × 5
  name      profession    birth_year death_year known_for
  <chr>     <chr>          <dbl>        <dbl>     <chr>
1 Ada Lovelace Mathematician      NA           NA   first co...
2 Marie Curie  Physicist and Chemist  NA           NA   theory o...
3 Janaki Ammal Botanist           1897         1984  hybrid s...
4 Chien-Shiung Wu Physicist         1912         1997  confirm a...
5 Katherine Johnson Mathematician    1918         2020  calculat...
6 Rosalind Franklin Chemist         1920         1958  <NA>
7 Vera Rubin    Astronomer        1928         2016  existenc...
8 Gladys West   Mathematician      1930         NA   mathemat...
9 Flossie Wong-Staal Virologist and Molecular ... 1947         NA   first sc...
10 Jennifer Doudna Biochemist       1964         NA   one of t...
```

Inputs als Erinnerung

```
1 names(professions)
```

```
[1] "name"      "profession"
```

```
1 names(dates)
```

```
[1] "name"      "birth_year"  
"death_year"
```

```
1 names(works)
```

```
[1] "name"      "known_for"
```

```
1 nrow(professions)
```

```
[1] 10
```

```
1 nrow(dates)
```

```
[1] 8
```

```
1 nrow(works)
```

```
[1] 9
```

Dataframes zusammenfügen

```
1 etwas_join(x, y)
```

- `left_join()`: alle Reihen aus x
- `right_join()`: alle Reihen aus y
- `full_join()`: alle Reihen aus x und y
- ...

Beispiel

```
1 x <- tibble(  
2   id = c(1, 2, 3),  
3   value_x = c("x1", "x2", "  
4 )
```

```
1 x
```

```
# A tibble: 3 × 2  
  id    value_x  
  <dbl> <chr>  
1     1  x1  
2     2  x2  
3     3  x3
```

```
1 y <- tibble(  
2   id = c(1, 2, 4),  
3   value_y = c("y1", "y2", "  
4 )
```

```
1 y
```

```
# A tibble: 3 × 2  
  id    value_y  
  <dbl> <chr>  
1     1  y1  
2     2  y2  
3     4  y4
```

left_join()

left_join(x, y)

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4

```
1 left_join(x, y)
```

```
# A tibble: 3 × 3
  id value_x value_y
  <dbl> <chr>   <chr>
1     1 x1      y1
2     2 x2      y2
3     3 x3      <NA>
```

left_join()

```
1 professions |>  
2   left_join(dates)
```

name	profession	birth_year	death_year
Ada Lovelace	Mathematician	NA	NA
Marie Curie	Physicist and Chemist	NA	NA
Janaki Ammal	Botanist	1897	1984
Chien-Shiung Wu	Physicist	1912	1997
Katherine Johnson	Mathematician	1918	2020
Rosalind Franklin	Chemist	1920	1958
Vera Rubin	Astronomer	1928	2016
Gladys West	Mathematician	1930	NA
Flossie Wong-Staal	Virologist and Molecular Biologist	1947	NA
Jennifer Doudna	Biochemist	1964	NA

right_join()

right_join(x, y)

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4

```
1 right_join(x, y)
```

```
# A tibble: 3 × 3
  id value_x value_y
  <dbl> <chr>   <chr>
1     1 x1      y1
2     2 x2      y2
3     4 <NA>    y4
```

right_join()

```
1 professions |>  
2   right_join(dates)
```

name	profession	birth_year	death_year
Janaki Ammal	Botanist	1897	1984
Chien-Shiung Wu	Physicist	1912	1997
Katherine Johnson	Mathematician	1918	2020
Rosalind Franklin	Chemist	1920	1958
Vera Rubin	Astronomer	1928	2016
Gladys West	Mathematician	1930	NA
Flossie Wong-Staal	Virologist and Molecular Biologist	1947	NA
Jennifer Doudna	Biochemist	1964	NA

full_join()

full_join(x, y)

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4

```
1 full_join(x, y)
```

```
# A tibble: 4 × 3
  id value_x value_y
  <dbl> <chr>   <chr>
1     1 x1      y1
2     2 x2      y2
3     3 x3      <NA>
4     4 <NA>    y4
```

full_join()

```
1 dates |>  
2 full_join(works)
```

name	birth_year	death_year	known_for
Janaki Ammal	1897	1984	hybrid species, biodiversity protection
Chien-Shiung Wu	1912	1997	confirm and refine theory of radioactive beta decay, Wu experiment overturning theory of parity
Katherine Johnson	1918	2020	calculations of orbital mechanics critical to sending the first Americans into space
Rosalind Franklin	1920	1958	NA
Vera Rubin	1928	2016	existence of dark matter
Gladys West	1930	NA	mathematical modeling of the shape of the Earth which served as the foundation of GPS technology
Flossie Wong-Staal	1947	NA	first scientist to clone HIV and create a map of its genes which led to a test for the virus
Jennifer Doudna	1964	NA	one of the primary developers of CRISPR, a ground-breaking technology for editing genomes
Ada Lovelace	NA	NA	first computer algorithm
Marie Curie	NA	NA	theory of radioactivity, discovery of elements polonium and radium, first woman to win a Nobel Prize

Alles in einer Code Sequenz

```
1 professions |>
2   left_join(dates) |>
3   left_join(works)
```

name	profession	birth_year	death_year	known_for
Ada Lovelace	Mathematician	NA	NA	first computer algorithm
Marie Curie	Physicist and Chemist	NA	NA	theory of radioactivity, discovery of elements polonium and radium, first woman to win a Nobel Prize
Janaki Ammal	Botanist	1897	1984	hybrid species, biodiversity protection
Chien-Shiung Wu	Physicist	1912	1997	confirm and refine theory of radioactive beta decay, Wu experiment overturning theory of parity
Katherine Johnson	Mathematician	1918	2020	calculations of orbital mechanics critical to sending the first Americans into space
Rosalind Franklin	Chemist	1920	1958	NA
Vera Rubin	Astronomer	1928	2016	existence of dark matter
Gladys West	Mathematician	1930	NA	mathematical modeling of the shape of the Earth which served as the foundation of GPS technology
Flossie Wong-Staal	Virologist and Molecular Biologist	1947	NA	first scientist to clone HIV and create a map of its genes which led to a test for the virus
Jennifer Doudna	Biochemist	1964	NA	one of the primary developers of CRISPR, a ground-breaking technology for editing genomes

Fehlende Daten ergänzen

```
1 professions |>
2   left_join(dates) |>
3   mutate(birth_year = case_when(
4     name == "Ada Lovelace" ~ 1815,
5     name == "Marie Curie" ~ 1867,
6     .default = birth_year
7   ))
```

A tibble: 10 × 4

	name	profession	birth_year	death_year
	<chr>	<chr>	<dbl>	<dbl>
1	Ada Lovelace	Mathematician	1815	NA
2	Marie Curie	Physicist and Chemist	1867	NA
3	Janaki Ammal	Botanist	1897	1984
4	Chien-Shiung Wu	Physicist	1912	1997
5	Katherine Johnson	Mathematician	1918	2020
6	Rosalind Franklin	Chemist	1920	1958
7	Vera Rubin	Astronomer	1928	2016
# i 3 more rows				

Ihr seid dran: 02-joins-ihr.qmd

1. Öffne [posit.cloud](#) in deinem Browser (verwende dein Lesezeichen).
2. Öffne den rstatszh-k010 Arbeitsbereich (Workspace) für den Kurs.
3. Klicke auf Start neben md-07-uebungen.
4. Suche im Dateimanager im Fenster unten rechts die Datei 02-joins-ihr.qmd und klicke darauf, um sie im Fenster oben links zu öffnen.
5. Folge den Anweisungen in der Datei.

Pause machen

Bitte steh auf und beweg dich. Lasst eure E-Mails in Frieden ruhen.

Ihr seid dran: 02-joins-ihr.qmd

1. Öffne [posit.cloud](#) in deinem Browser (verwende dein Lesezeichen).
2. Öffne den rstatszh-k010 Arbeitsbereich (Workspace) für den Kurs.
3. Klicke auf Start neben md-07-uebungen.
4. Suche im Dateimanager im Fenster unten rechts die Datei 02-joins-ihr.qmd und klicke darauf, um sie im Fenster oben links zu öffnen.
5. Folge den Anweisungen in der Datei.

Zeitpuffer: Modul 7

Kann ich noch etwas zum
heutigen Modul erklären?

Zusatzaufgaben Modul 7

Modul 7 Dokumentation

rstatszh-k010.github.io/website/module/md-07.html

Zusatzaufgaben Abgabedatum

- Abgabedatum: Montag, 14. April
- Korrektur- und Feedbackphase bis zu: Donnerstag, 17. April

Danke

Danke! 🌻

Folien erstellt mit revealjs und Quarto:

<https://quarto.org/docs/presentations/revealjs/>

Access slides als [PDF auf GitHub](#)

Alle Materialien sind lizenziert unter [Creative Commons Attribution Share Alike 4.0 International.](#)