

Präsentation der Auswertungsstrategie

Datensatz Olympische Geschichte

Felix Vogel

01.03.2022

Übersicht über den Datensatz

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
1	A Dijiang	M	24	180	80	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NA
2	A Lamusi	M	23	170	60	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NA
3	Gunnar Nielsen Aaby	M	24	NA	NA	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NA
4	Edgar Lindenau Aabye	M	34	NA	NA	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NA
5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 1,000 metres	NA
5	Christine Jacoba Aaftink	F	25	185	82	Netherlands	NED	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Women's 500 metres	NA
5	Christine Jacoba Aaftink	F	25	185	82	Netherlands	NED	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Women's 1,000 metres	NA
5	Christine Jacoba Aaftink	F	27	185	82	Netherlands	NED	1994 Winter	1994	Winter	Lillehammer	Speed Skating	Speed Skating Women's 500 metres	NA
5	Christine Jacoba Aaftink	F	27	185	82	Netherlands	NED	1994 Winter	1994	Winter	Lillehammer	Speed Skating	Speed Skating Women's 1,000 metres	NA

Auswertungsstrategie

- Deskriptive Statistiken

zum Beispiel: "ewiger Medaillenspiegel"

→ dazu "lexikographische" Ordnung der Länder nach ihren Medaillengewinnen in der Rangfolge Gold, Silber, Bronze

```
#Left Join der NOC-Codes zu olympia
olympia_mit_ländern <- olympia %>%
  left_join(NOC_Codes, by="NOC")

Medaillen<-olympia_mit_ländern %>%
  group_by(region, Medal) %>%
  na.omit() %>%
  summarize(Anzahl=n()) %>%
  arrange(desc(Anzahl))
Medaillen_wide <- pivot_wider(Medaillen, names_from = "Medal", values_from = "Anzahl")
```

Auswertungsstrategie

region	Gold	Silver	Bronze
USA	2115	1286	982
Russia	1465	1062	1083
Germany	1102	1015	1072
Australia	319	414	477
Italy	315	358	387
France	285	327	375

→ Sortierung (Gold, Silber, Bronze) wahrscheinlich "zufällig" korrekt

- weitere Ideen z.B. Karte mit Austragungsorten, deskriptive Athletendaten

Auswertungsstrategie

- Regressionsanalyse: "Was unterscheidet Medaillengewinner?"

Hierarchische Regression: "offensichtliche" Hierarchie nach: Saison, Sportart, Wettkampf

Season	Sport	Anzahl_Events
Summer	Athletics	83
Summer	Shooting	83
Summer	Swimming	55
Summer	Cycling	44
Summer	Sailing	38
Summer	Wrestling	30
Summer	Archery	29
Summer	Art Competitions	29
Summer	Canoeing	27

- Regressionsanalyse
- zusätzliche unabhängige Variablen: Größe, Alter, Gewicht, Land
- Untersuchung getrennt nach Männern und Frauen
- abhängige Variable: Medaille (Gold, Silber, Bronze, Keine)
- hierarchische Regression führt Variablen schrittweise, aber nach Kausalität ein
→ d.h. kausal grundlegendere Variablen zuerst, danach ggf. von diesen beeinflusste

Für diese Hierarchie ergibt sich dann:

- 1. Block: Größe, Alter, Gewicht (metrisch)
- danach Season, Sport, Event (nominal)
- zuletzt Land (nominal)