

## Beispieltitel

unbekannte:r Autor:in

unbekanntes Datum

Im Folgenden nutzen wir die bisher erstellten Graphiken, Tabellen und Kennzahlen, um einen HTML-Report für die Break Free From Plastic Bewegung zu erstellen. Das Layout von HTML-Reports ist sehr zugänglich und die Berichte können über einen Doppelklick im Browser geöffnet werden.

## Übung 1: Markdown (Inhalt)

Nutzt die gelernten Formatierungsoptionen, um im Folgenden die Beschreibung des Plastics-Datensatz zu verbessern.

Titel: Eine explorative Datenanalyse der Break Free From Plastic Bewegung für das Jahr 2019

Titelbild (einbetten): <https://camo.githubusercontent.com/e571cad1e53d23d5a8fbc452989601fdd98de550e5ce165762a8cc290f68747470733a2f2f7777772e627265616b6672656566726f6d706c61737469632e6f72672f77702d636f6e74656e742f75706c6f6164732f>

Im Folgenden soll der Datensatz der Break Free From Plastic (fett) Bewegung für das Jahr 2019 und 2020 explorativ erkundet werden. Der Datensatz enthält 13.380 Beobachtungen zu Plastiksammelaktionen in 69 Ländern. Enthalten sind die Variablen (als Stichpunkte): Land, Jahr, Hersteller des Produktes, Art des gesammelten Plastiks sowie die Anzahl an durchgeführten Events und der Anzahl teilnehmender Freiwilliger.

Anmerkung (kursiv): Mehr Informationen zu Break Free From Plastic finden Sie unter diesem Link (als Hyperlink einfügen, <https://www.breakfreefromplastic.org/>).

## Schritt 1: Infrastruktur

## Schritt 2: Daten laden

```
#### Daten laden
plastics <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/
## Rows: 13380 Columns: 14

## -- Column specification -----
## Delimiter: ","
## chr (2): country, parent_company
## dbl (12): year, empty, hdpe, ldpe, o, pet, pp, ps, pvc, grand_total, num_eve...

##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

## Schritt 3: Daten bereinigen

```
### Daten bereinigen
# Country name cleaning:
# United Kingdom of Great Britain & Northern Ireland sowie für die USA
plastics_prep <- plastics %>%
  # nur 2019 weil 2020 wegen der Pandemie ein nicht repräsentatives Jahr ist
  filter(year==2019) %>%
  mutate(country = str_replace(country, "United Kingdom of Great Britain & Northern Ireland", "United K
        country = str_replace(country, "United States of America", "United States"),
        country = str_to_title(country) ) %>% #um z.B.: ECUADOR etc case ändern
  #Continent und Country Code anspielen
  mutate(continent = countrycode::countrycode(country, origin = "country.name", destination = "continent
        countrycode = countrycode::countrycode(country, origin = "country.name", destination = "iso3c"
  mutate(continent=replace_na(continent, "Unknown"),
        countrycode=replace_na(countrycode, "Unknown"))
```

```
## Warning in countrycode_convert(sourcevar = sourcevar, origin = origin, destination = dest, : Some va
```

```
## Warning in countrycode_convert(sourcevar = sourcevar, origin = origin, destination = dest, : Some va
```

```
### Community Datensatz erstellen
community<- plastics_prep %>%
  select(country, year, num_events, volunteers, grand_total, continent, countrycode) %>%
  group_by(country, year) %>%
  filter(row_number()==1)

# Audit Plastik Datensatz erstellen
audit_plastic<- plastics_prep %>%
  # nur grandtotal zeile behalten
  filter(parent_company=="Grand Total") %>%
  select(-c(parent_company, num_events, volunteers, empty)) %>%
  #NA zu 0
  mutate(
    across(everything(), ~replace_na(.x, 0))
  )

# EU-Asien Audit Plastik Datensatz erstellen
audit_plastic_eu_asia<- audit_plastic %>%
  filter(continent == "Europe" | continent == "Asia") %>%
  group_by(country) %>%
  mutate(n_types=sum(c(hdpe, ldpe, o, pet, pp, ps, pvc)!=0))
```

## Schritt 4: Übersicht verschaffen

```
# Überblick über den Datensatz verschaffen
dplyr::glimpse(plastics_prep)
```

```
## Rows: 9,296
```

```
## Columns: 16
## $ country      <chr> "Argentina", "Argentina", "Argentina", "Argentina", "Ar~
## $ year         <dbl> 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2~
## $ parent_company <chr> "Grand Total", "Unbranded", "The Coca-Cola Company", "S~
## $ empty        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ hdpe         <dbl> 215, 155, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ ldpe         <dbl> 55, 50, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ o            <dbl> 607, 532, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 13, 0, 0, ~
## $ pet          <dbl> 1376, 848, 222, 39, 38, 22, 21, 26, 19, 14, 14, 14, 14, ~
## $ pp           <dbl> 281, 122, 35, 4, 0, 7, 6, 0, 1, 4, 3, 1, 0, 0, 3, 0, 4, ~
## $ ps           <dbl> 116, 114, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ pvc          <dbl> 18, 17, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ grand_total  <dbl> 2668, 1838, 257, 43, 38, 29, 27, 26, 20, 18, 17, 15, 14~
## $ num_events   <dbl> 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4~
## $ volunteers   <dbl> 243, 243, 243, 243, 243, 243, 243, 243, 243, 243, 243, ~
## $ continent    <chr> "Americas", "Americas", "Americas", "Americas", "Americ~
## $ countrycode  <chr> "ARG", "ARG", "ARG", "ARG", "ARG", "ARG", "ARG", "ARG", ~
```

```
# Eigenschaften des Datensatzes
nrow <- nrow(plastics_prep)
laender <- length(unique(plastics_prep$country))
```

## Übung 2: Code Chunks

Wir haben bis zu diesem Schritt einige Code Chunks aus den vergangenen Übungen kopiert. Manche von ihnen werfen Warnmeldungen und Nachrichten zurück. Diese sollen für das Outputformat grundsätzlich unterdrückt werden. Definiert im Set-Up-Chunk die Argumente so, dass das der Fall ist.

## Übung 3: Code Chunks

Die Schritte 2-4 dienen in der Regel nur uns - im späteren Report sollen sie nicht enthalten sein. Setzt die Argumente in den individuellen Code Chunks so, dass sie nicht enthalten sind.

## Übung 4: Inline Code

Wir möchten den Leser:innen mitteilen, dass wir uns während der Datenbereinigung dafür entschieden haben, das Jahr 2020 nicht zu berücksichtigen, da dort weniger Events stattfinden konnten. Dazu gibt es zwei Optionen: 1. Wir formulieren die neue Anzahl an Beobachtungen (9296) aus. 2. Wir greifen im Text auf die berechnete Variable zu.

Beispielsweise können wir über 9296 die Anzahl an Beobachtungen in den Text einfügen. Schaut Euch das gerne im PDF-Dokument erstmal an.

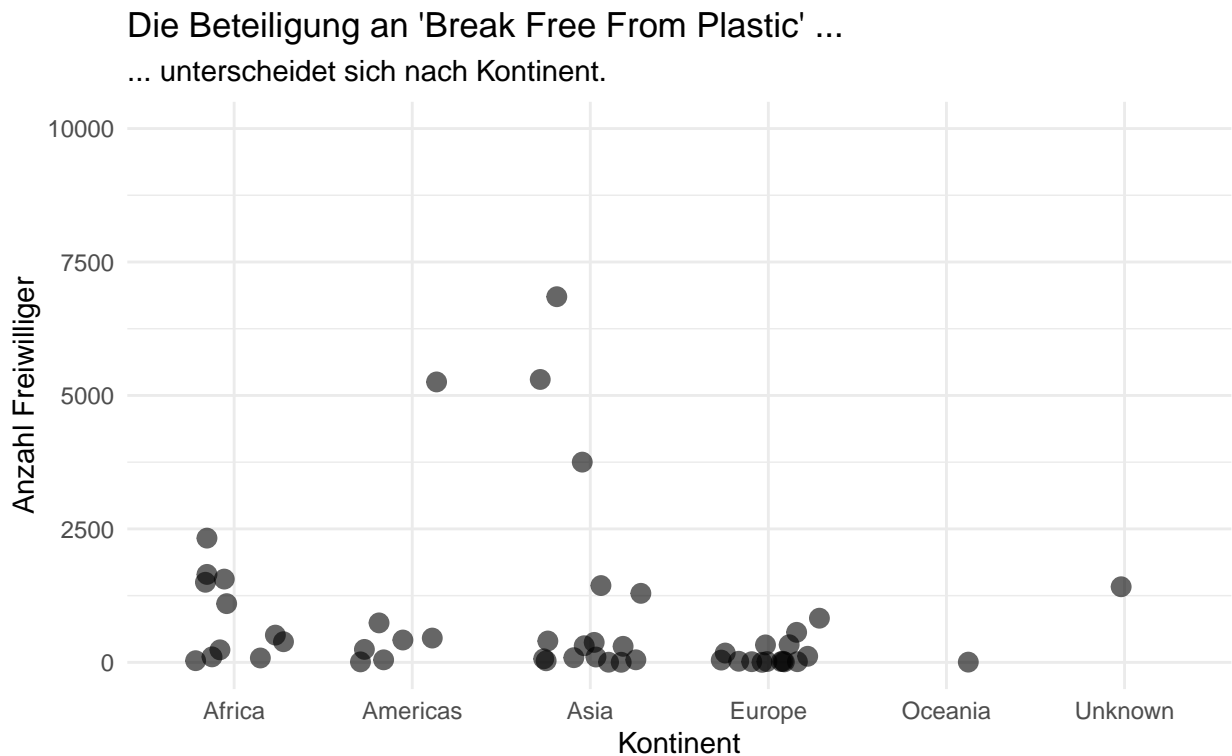
## Schritt 5: Visuelle Exploration

```
# Erstellung eines Scatterplots zu der Anzahl an Freiwilligen
ggplot(data = community, aes(x = continent, y = volunteers)) + # Initialisierung des ggplots mit Variab
  geom_point(position = position_jitter(width = 0.3),
             size = 3,
```

```

    alpha = 0.6) + # Hinzufügen der Datenpunkte (Scatterplot) inkl. Stylingoptionen zur Positionierung
coord_cartesian(ylim = c(0, 10000)) + # Festlegung der Achsenlänge der y-Achse
labs(
  title = "Die Beteiligung an 'Break Free From Plastic' ..." ,
  subtitle = "... unterscheidet sich nach Kontinent.",
  y = "Anzahl Freiwilliger",
  x = "Kontinent",
  caption = "In Taiwan haben sich 31.318 Freiwillige beteiligt. Diese Beobachtung \nwurde zur Lesbarkeit
) + # Festlegung der Achsenbezeichnungen, Überschriften und Titel
theme_minimal() # Festlegung des Layout-Designs

```



In Taiwan haben sich 31.318 Freiwillige beteiligt. Diese Beobachtung wurde zur Lesbarkeit des Graphen ausgeklammert.  
Datenquelle: TidyTuesday und BFFP

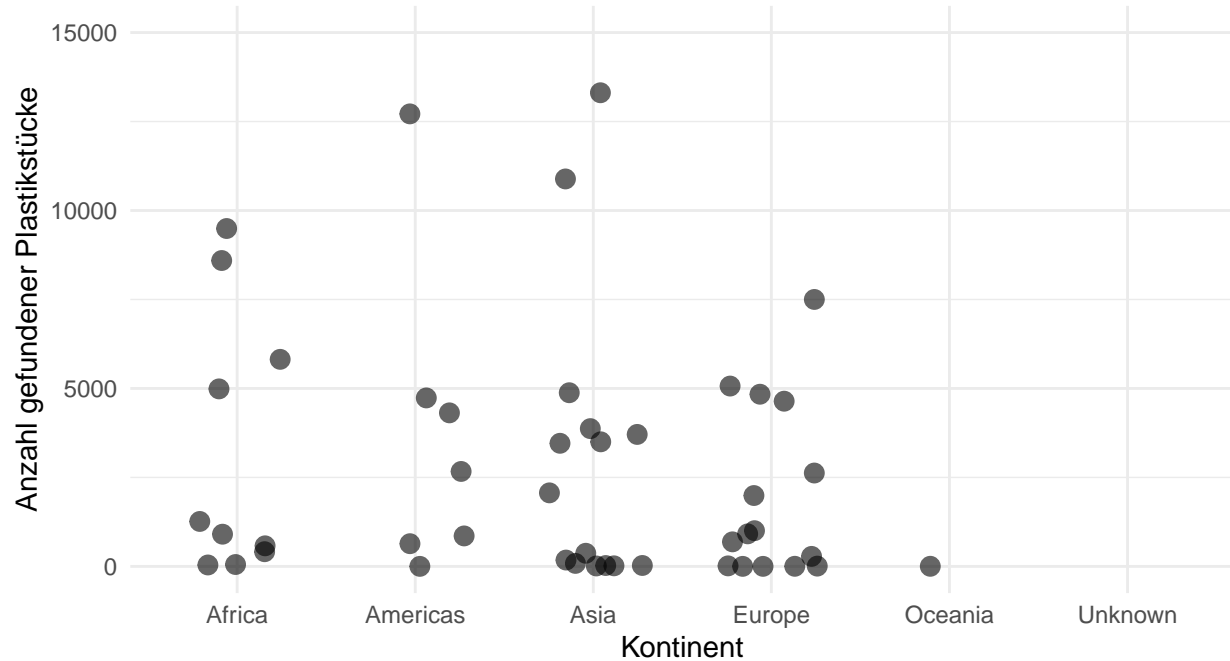
```

# Erstellung eines Punktediagramms mit der Anzahl gesammelter Plastikstücke pro Kontinent
ggplot(data = audit_plastic, aes(continent, grand_total)) + # Initialisierung des ggplots mit Variablen
  geom_point(position = position_jitter(width = 0.3),
    size = 3,
    alpha = 0.6) + # Hinzufügen der Datenpunkte (Scatterplot) inkl. Stylingoptionen zur Positionierung
coord_cartesian(ylim = c(0, 15000)) + # Festlegung der Achsenlänge der y-Achse
labs(
  title = "Beteiligung an 'Break Free From Plastic' ..." ,
  subtitle = "... unterscheidet sich nach Kontinent.",
  y = "Anzahl gefundener Plastikstücke",
  x = "Kontinent",
  caption = "In Nigeria, Philippinen und Taiwan wurden 19.841, 28.055 und \n120.646 Plastikstücke gesammelt.
) + # Festlegung der Achsenbezeichnungen, Überschriften und Titel
theme_minimal() # Festlegung des Layout-Designs

```

## Beteiligung an 'Break Free From Plastic' ...

... unterscheidet sich nach Kontinent.



In Nigeria, Philippinen und Taiwan wurden 19.841, 28.055 und 120.646 Plastikstücke gesammelt. Diese Beobachtungen wurden zur Lesbarkeit ausgeklammert.

## Schritt 6: Statistische Kennzahlen

```
# Berechnung statistischer Kennzahlen pro Kontinent
audit_plastic %>%
  group_by(continent) %>%
  summarize(
    # Mittelwert
    menge_mittelwert = mean(grand_total),
    # Standardabweichung
    menge_standardabweichung = sd(grand_total),
    # Median
    menge_median = median(grand_total),
    # Anzahl beteiligter Länder
    länder_anzahl = n(),
    # Summe der Plastikmenge
    menge_summe = sum(grand_total)
  )
```

continent	menge_mittelwert	menge_standardabweichung	menge_median	länder_anzahl	menge_summe
Africa	10247.000	23584.267	1263	11	112717
Americas	3703.714	4376.606	2668	7	25926
Asia	12003.000	29424.991	3459	17	204051

continent	menge_mittelwert	menge_standardabweichung	menge_median	länder_anzahl	menge_summe
Europe	1971.933	2417.334	914	15	29579
Oceania	3.000	NA	3	1	3
Unknown	56955.000	NA	56955	1	56955

```
# Berechnung der Länder- und Freiwilligenanzahl je Kontinent
community %>% # Bezug zum Datensatz
  group_by(continent) %>% # Gruppierung pro Kontinent
  summarize( # Initialisierung der Berechnung
    # Anzahl beteiligter Länder
    länder_anzahl = n(),
    # Teilnahmezahl der Freiwilligen
    freiwillige_anzahl = sum(volunteers))
```

continent	länder_anzahl	freiwillige_anzahl
Africa	11	9486
Americas	7	7167
Asia	17	51675
Europe	15	2487
Oceania	1	5
Unknown	1	1416

Nina: Inline Code - Kennzahl einfügen

```
head(plastics)
```

country	year	parent_company	empty	hdpe	ldpe	o	pet	pp	ps	pvc	grand_total	num_events	volunteers
Argentina	2019	Grand Total	0	215	55	607	1376	281	116	18	2668	4	243
Argentina	2019	Unbranded	0	155	50	532	848	122	114	17	1838	4	243
Argentina	2019	The Coca-Cola Company	0	0	0	0	222	35	0	0	257	4	243
Argentina	2019	Secco	0	0	0	0	39	4	0	0	43	4	243
Argentina	2019	Doble Cola	0	0	0	0	38	0	0	0	38	4	243
Argentina	2019	Pritty	0	0	0	0	22	7	0	0	29	4	243

```
#TODO
```

### Schritt 3: YAML Kopfzeile