Covid-19 Vaccination Campaign in Germany

The data used here were provided by Robert Koch Institute and the German federal ministry of Health.

These institutions publish the datasets and some analysis on the page impfdashboard.de.

Setup

Imports

In [66]:
```
# standard library
import datetime
import math
```

In [67]:
```
# third party
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import requests
import seaborn
```

Date this Notebook was run

In [68]:
```
today = datetime.datetime.today().strftime('%Y-%m-%d')
today
```

Out[68]:
```
'2021-05-19'
```

Set Defaults

In [69]:
```
# style like ggplot in R
plt.style.use('ggplot')
```

In [70]:
```
# Avoid cutting off part of the axis labels, see:
# https://stackoverflow.com/questions/6774086/why-is-my-xlabel-cut-off-in-my-matplotlib-plot
plt.rcParams.update({'figure.autolayout': True})
```

In [71]:
```
population_germany = 83_200_000
```

Get and Transform Data

In [72]:
```
vaccination_data_permalink = 'https://impfdashboard.de/static/data/germany_vaccinations_timeseries_v2.tsv'
vaccinations = pd.read_csv(
    vaccination_data_permalink,
    sep="\t")
```

Drop unnecessary / misleading columns

Columns with names starting with 'indikation_' will not be analyzed as the data providers stopped updating them.

In [73]:
```
cols_to_drop = vaccinations.columns[vaccinations.columns.str.contains('indikation_')]
vaccinations.drop(columns=cols_to_drop, inplace=True)
```

Some more columns can be dropped, as there is no interest in analyzing differences on a vaccine level - especially since in some cases vaccines were mixed.

In [74]:
```
more_cols_to_drop = ['dosen_biontech_erst_kumulativ', 'dosen_biontech_zweit_kumulativ',
                     'dosen_moderna_erst_kumulativ', 'dosen_moderna_zweit_kumulativ',
                     'dosen_astrazeneca_erst_kumulativ', 'dosen_astrazeneca_zweit_kumulativ']
vaccinations.drop(columns=more_cols_to_drop, inplace=True)
```

Some columns are labeled misleadingly. As stated by the data provider the columns `personen_erst_kumulativ` and `impf_quote_erst` contain people vaccinated with the Johnson & Johnson vaccine. As this requires only one shot. the same persons are included in `personen_voll_kumulativ`. Therefore more columns are dropped and recalculated later.

In [75]:
```
vaccinations.drop(columns=['impf_quote_erst', 'impf_quote_voll'], inplace=True)
```

Convert datatype of date column

In [76]:
```
vaccinations.iloc[ : , [0]] = vaccinations.iloc[ : , [0]].apply(pd.to_datetime)
```

Show Data

In [77]:
```
vaccinations.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 143 entries, 0 to 142
Data columns (total 13 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   date                            143 non-null    datetime64[ns]
 1   dosen_kumulativ                 143 non-null    int64
 2   dosen_differenz_zum_vortag      143 non-null    int64
 3   dosen_erst_differenz_zum_vortag 143 non-null    int64
 4   dosen_zweit_differenz_zum_vortag 143 non-null   int64
 5   dosen_biontech_kumulativ        143 non-null    int64
 6   dosen_moderna_kumulativ         143 non-null    int64
 7   dosen_astrazeneca_kumulativ     143 non-null    int64
 8   personen_erst_kumulativ         143 non-null    int64
 9   personen_voll_kumulativ         143 non-null    int64
 10  dosen_dim_kumulativ             143 non-null    int64
 11  dosen_kbv_kumulativ             143 non-null    int64
 12  dosen_johnson_kumulativ         143 non-null    int64
dtypes: datetime64[ns](1), int64(12)
memory usage: 14.6 KB
```

In [78]:
```
vaccinations.tail(3)
```

Out[78]:

	date	dosen_kumulativ	dosen_differenz_zum_vortag	dosen_erst_differenz_zum_vortag	dosen_zweit_differenz_zum_vortag	dosen_biontech_kumulati
140	2021-05-16	40141267	274659	157754	116905	2960938
141	2021-05-17	40673871	532604	341985	190619	2993614
142	2021-05-18	41517849	843978	508111	335867	3051116

Check Validity

In [79]:
```
# get the last row / the newest available data
last_row = vaccinations.tail(1)
```

In [80]:
```
doses_used = last_row['dosen_kumulativ']
doses_used
```

Out[80]:
```
142    41517849
Name: dosen_kumulativ, dtype: int64
```

In [81]:
```
# The number of person having been vaccinated at least once, includes those fully vaccinated
at_least_once = last_row['personen_erst_kumulativ']
fully_vaccinated_people = last_row['personen_voll_kumulativ']
partially_vaccinated_people = at_least_once - fully_vaccinated_people
# The johnson & Johnson vaccine is the only one used in Germany that only needs a single shot:
johnson_doses = last_row['dosen_johnson_kumulativ']
```

In [82]:
```
# Must be exactly 0
doses_used - partially_vaccinated_people - (fully_vaccinated_people - johnson_doses) * 2 - johnson_doses ==  0
```

Out[82]:
```
142    True
dtype: bool
```

Calculate columns

In [83]:
```
vaccinations['partly vaccinated'] = round(
    (vaccinations['personen_erst_kumulativ'] - vaccinations['personen_voll_kumulativ']) * 100 / population_germany,
    2)
```

In [84]:
```
vaccinations['fully vaccinated'] = round(
    vaccinations['personen_voll_kumulativ'] * 100 / population_germany,
    2)
```

In [85]:
```
vaccinations.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 143 entries, 0 to 142
Data columns (total 15 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   date                            143 non-null    datetime64[ns]
 1   dosen_kumulativ                 143 non-null    int64
 2   dosen_differenz_zum_vortag      143 non-null    int64
 3   dosen_erst_differenz_zum_vortag 143 non-null    int64
 4   dosen_zweit_differenz_zum_vortag 143 non-null   int64
 5   dosen_biontech_kumulativ        143 non-null    int64
 6   dosen_moderna_kumulativ         143 non-null    int64
 7   dosen_astrazeneca_kumulativ     143 non-null    int64
 8   personen_erst_kumulativ         143 non-null    int64
 9   personen_voll_kumulativ         143 non-null    int64
 10  dosen_dim_kumulativ             143 non-null    int64
 11  dosen_kbv_kumulativ             143 non-null    int64
 12  dosen_johnson_kumulativ         143 non-null    int64
 13  partly vaccinated               143 non-null    float64
```

```
 14  fully vaccinated                  143 non-null    float64
dtypes: datetime64[ns](1), float64(2), int64(12)
memory usage: 16.9 KB
```

In [86]:
```
vaccinations.tail(3)
```

Out[86]:

	date	dosen_kumulativ	dosen_differenz_zum_vortag	dosen_erst_differenz_zum_vortag	dosen_zweit_differenz_zum_vortag	dosen_biontech_kumulati
140	2021-05-16	40141267	274659	157754	116905	2960938
141	2021-05-17	40673871	532604	341985	190619	2993614
142	2021-05-18	41517849	843978	508111	335867	3051116

Last Update

Often the data is not updated on weekends, so get the highest date in the dataset.

In [87]:
```
last_update = vaccinations.loc[vaccinations.index[-1], "date"].strftime('%Y-%m-%d')
last_update
```

Out[87]:
```
'2021-05-18'
```

Doses Used

In [88]:
```
doses = vaccinations.loc[ : , ['date', 'dosen_differenz_zum_vortag']]
# Rename columns
doses.columns = ['date', 'doses used']
```

In [89]:
```
# Scale number of doses as millions
doses['doses used'] = doses['doses used'] / 1_000_000
```

Doses Daily

In [90]:
```
doses_daily = doses.set_index('date', inplace=False)
doses_daily.tail(1)
```

Out[90]:

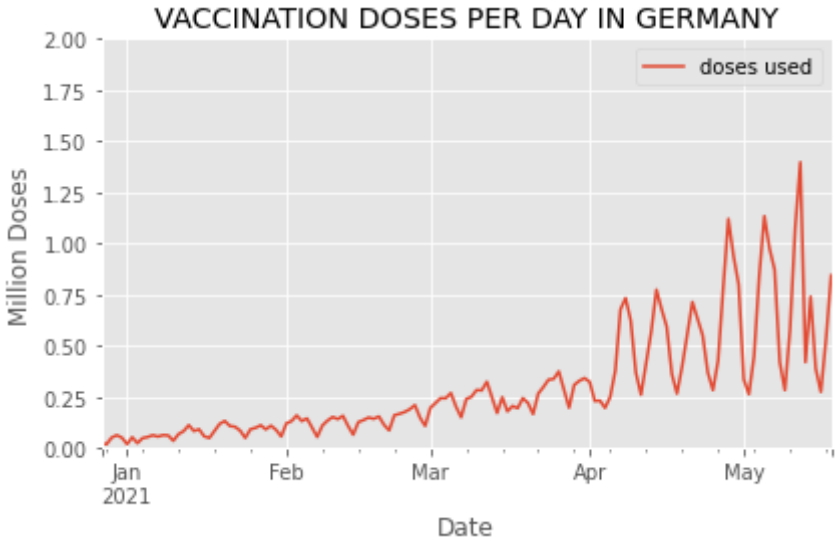
date	doses used
2021-05-18	0.843978

In [91]:
```
# What is the highest number of doses used in a day?
max_doses_daily = max(doses_daily['doses used'])
max_doses_daily
```

Out[91]: 1.397345

In [92]:
```
doses_daily.plot(
    ylim=(0,math.ceil(max_doses_daily)),
    xlabel='Date',
    ylabel='Million Doses',
    title='VACCINATION DOSES PER DAY IN GERMANY')
```

Out[92]: <AxesSubplot:title={'center':'VACCINATION DOSES PER DAY IN GERMANY'}, xlabel='Date', ylabel='Million Doses'>



Doses per Weekday (in the last 6 weeks)

In [93]:
```
last_6_weeks = doses.tail(42)
```

In [94]:
```
# Yields a warning, but exactly like the docs prescribe and it works
# https://pandas.pydata.org/docs/getting_started/intro_tutorials/05_add_columns.html
last_6_weeks['weekday'] = last_6_weeks['date'].dt.day_name()
```

```
<ipython-input-94-45013977109e>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  last_6_weeks['weekday'] = last_6_weeks['date'].dt.day_name()
```

In [95]:
```
# check:
last_6_weeks.tail(3)
```

Out[95]:

	date	doses used	weekday
140	2021-05-16	0.274659	Sunday
141	2021-05-17	0.532604	Monday
142	2021-05-18	0.843978	Tuesday

In [96]:
```
# drop the date column
last_6_weeks = last_6_weeks.drop(labels=['date'], axis=1)
```

In [97]:
```
#last_6_weeks.set_index('weekday', inplace=True)
last_6_weeks.tail(3)
```

Out[97]:

	doses used	weekday
140	0.274659	Sunday
141	0.532604	Monday
142	0.843978	Tuesday

In [98]:
```
pivot_table =last_6_weeks.pivot(columns='weekday', values='doses used')
pivot_table.tail()
```

Out[98]:

weekday	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
138	0.740226	NaN	NaN	NaN	NaN	NaN	NaN
139	NaN	NaN	0.38833	NaN	NaN	NaN	NaN

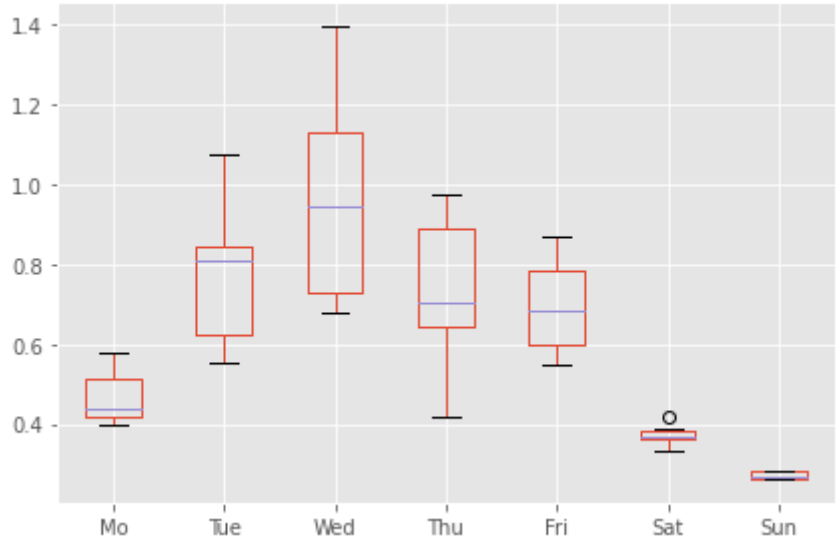
weekday	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
140	NaN	NaN	NaN	0.274659	NaN	NaN	NaN
141	NaN	0.532604	NaN	NaN	NaN	NaN	NaN
142	NaN	NaN	NaN	NaN	NaN	0.843978	NaN

In [99]:
```
# Reorder the columns
pivot_table = pivot_table[['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday']]
# Rename the columns
pivot_table.columns=['Mo', 'Tue', 'Wed', 'Thu', 'Fri', 'Sat', 'Sun']
pivot_table.tail()
```

Out[99]:

	Mo	Tue	Wed	Thu	Fri	Sat	Sun
138	NaN	NaN	NaN	NaN	0.740226	NaN	NaN
139	NaN	NaN	NaN	NaN	NaN	0.38833	NaN
140	NaN	NaN	NaN	NaN	NaN	NaN	0.274659
141	0.532604	NaN	NaN	NaN	NaN	NaN	NaN
142	NaN	0.843978	NaN	NaN	NaN	NaN	NaN

In [100…
```
weekday_boxplot = pivot_table.boxplot()
```



In [101…
```
fig = weekday_boxplot.get_figure()
fig.savefig('img/weekday_boxplot.png')
```

Doses per Week

In [102…
```
# W-Mon in order to start the week on a Monday, see:
# https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html#anchored-offsets
doses_weekly = doses.groupby(pd.Grouper(key='date',freq='W-Mon')).sum()
doses_weekly.columns = ['million doses used']
doses_weekly.tail()
```

Out[102…

date	million doses used
2021-04-26	3.534031
2021-05-03	4.690977
2021-05-10	5.105094
2021-05-17	4.827047
2021-05-24	0.843978

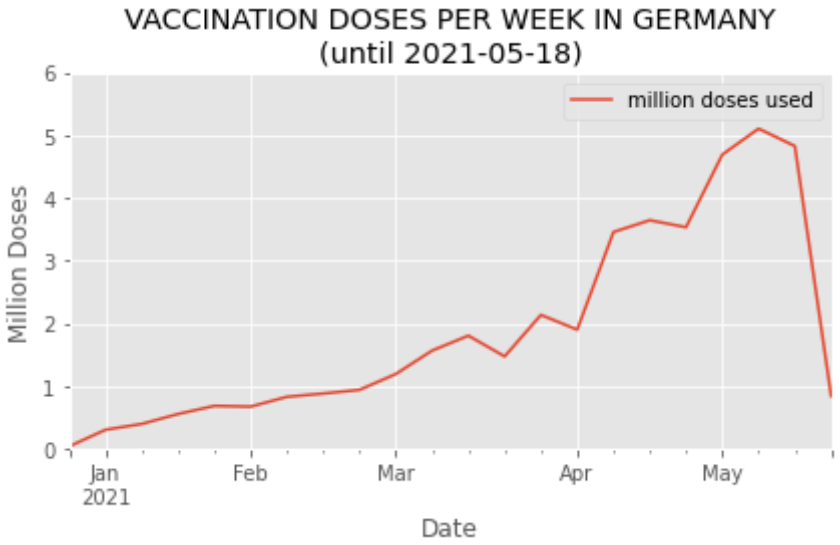
In [103…
```
# What is the highest number of doses used in a week?
```

```
max_million_doses_weekly = max(doses_weekly['million doses used'])
max_million_doses_weekly
```

Out[103… 5.105093999999999

In [104…
```
doses_weekly.plot(
    ylim=(0, math.ceil(max_million_doses_weekly)),
    xlabel='Date',
    ylabel='Million Doses',
    title=f"VACCINATION DOSES PER WEEK IN GERMANY\n(until {last_update})")
```

Out[104… <AxesSubplot:title={'center':'VACCINATION DOSES PER WEEK IN GERMANY\n(until 2021-05-18)'}, xlabel='Date', ylabel='Million Doses'>



Doses per Month

In [105…
```
# M = month end frequency
doses_monthly = doses.groupby(pd.Grouper(key='date',freq='M')).sum()
doses_monthly.tail()
```

Out[105…

	doses used
date	
2021-01-31	2.344521

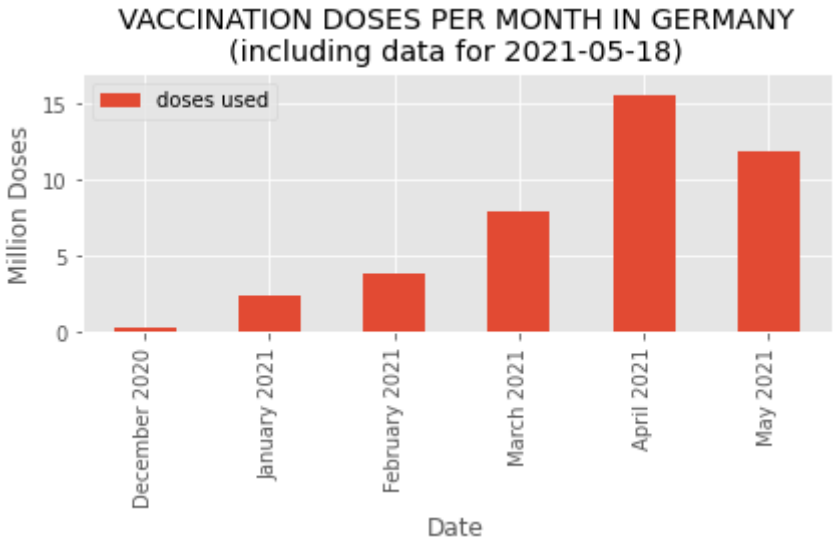
date	doses used
2021-02-28	3.778865
2021-03-31	7.852521
2021-04-30	15.511982
2021-05-31	11.823516

In [106…
```
max_doses_monthly = max(doses_monthly['doses used'])
max_doses_monthly
doses_monthly['month'] = doses_monthly.index.strftime('%B')
doses_monthly['year'] =  doses_monthly.index.strftime('%Y')
doses_monthly['label'] = doses_monthly['month'] + ' ' + doses_monthly['year']
doses_monthly.drop(columns=['month', 'year'], inplace=True)
doses_monthly.set_index('label', inplace=True)
doses_monthly.tail(6)
```

Out[106…

label	doses used
December 2020	0.206444
January 2021	2.344521
February 2021	3.778865
March 2021	7.852521
April 2021	15.511982
May 2021	11.823516

In [107…
```
monthly_plot = doses_monthly.plot.bar(
    ylim=(0,math.ceil(max_doses_monthly) + 1),
    xlabel='Date',
    ylabel='Million Doses',
    title=f"VACCINATION DOSES PER MONTH IN GERMANY\n(including data for {last_update})")
```



In [108…
```
fig = monthly_plot.get_figure()
fig.savefig('img/monthly_doses_germany.png')
```

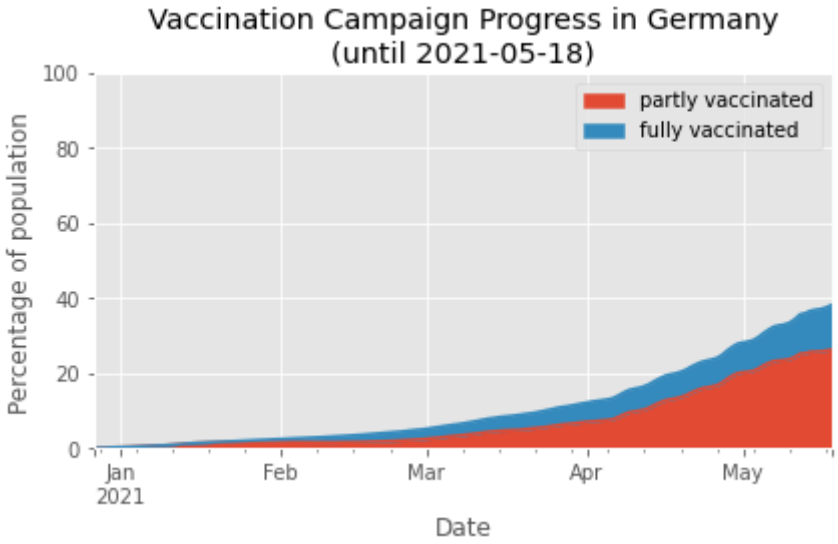
Vaccination Campaign Progress

In [109…
```
doses_cumulative = vaccinations.loc[ : , ['date', 'partly vaccinated', 'fully vaccinated']]
doses_cumulative.set_index('date', inplace=True)
doses_cumulative.tail(3)
```

Out[109…

date	partly vaccinated	fully vaccinated
2021-05-16	25.77	11.27
2021-05-17	25.96	11.50
2021-05-18	26.17	11.90

In [110…
```
doses_area_plot = doses_cumulative.plot.area(
    ylim=(0,100),
    xlabel='Date',
    ylabel='Percentage of population',
    title=f"Vaccination Campaign Progress in Germany\n(until {last_update})")
```



In [111…
```
fig = doses_area_plot.get_figure()
fig.savefig('img/vaccinations_germany_area_plot.png')
```

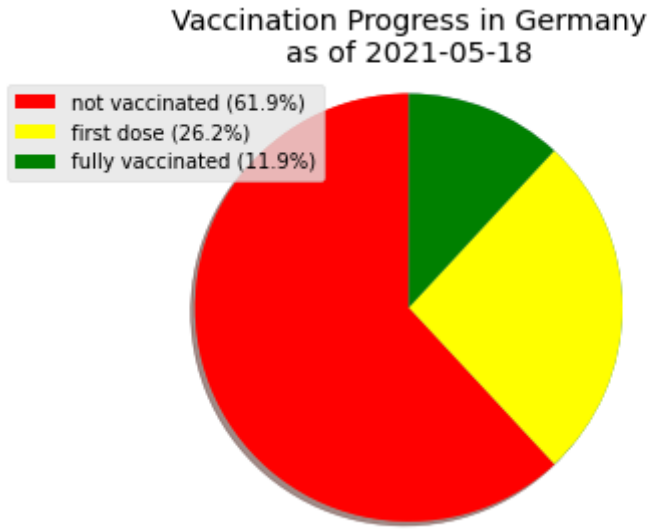
As of Today

In [112…
```
# get the last line of the data
current_state = doses_cumulative.iloc[-1]
current_state
```

Out[112…
```
partly vaccinated    26.17
fully vaccinated     11.90
Name: 2021-05-18 00:00:00, dtype: float64
```

In [113…
```
percentage_not_vacc = 100 - current_state['partly vaccinated'] - current_state['fully vaccinated']
labels = [f"not vaccinated ({round(percentage_not_vacc, 1)}%)",
          f"first dose ({round(current_state['partly vaccinated'], 1)}%)",
          f"fully vaccinated ({round(current_state['fully vaccinated'], 1)}%)"]
colors = ['red', 'yellow', 'green']
sizes = [percentage_not_vacc,
         current_state['partly vaccinated'],
         current_state['fully vaccinated']]
fig1, ax1 = plt.subplots()
ax1.pie(sizes, shadow=True, startangle=90)
ax1.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
patches, texts = plt.pie(sizes, colors=colors, startangle=90)
```

```
plt.legend(patches, labels, loc="best")
plt.title(f"Vaccination Progress in Germany\nas of {last_update}")
# plt.savefig must be before show()
# BEWARE plt.savefig must be in the same Jupyter code cell that creates the graph!
# See comment by ijoseph here:
# https://stackoverflow.com/questions/9012487/matplotlib-pyplot-savefig-outputs-blank-image
plt.savefig('img/vaccination_in_germany_pie.png', bbox_inches='tight')
plt.show()
```



Vaccines in Use

In [114…

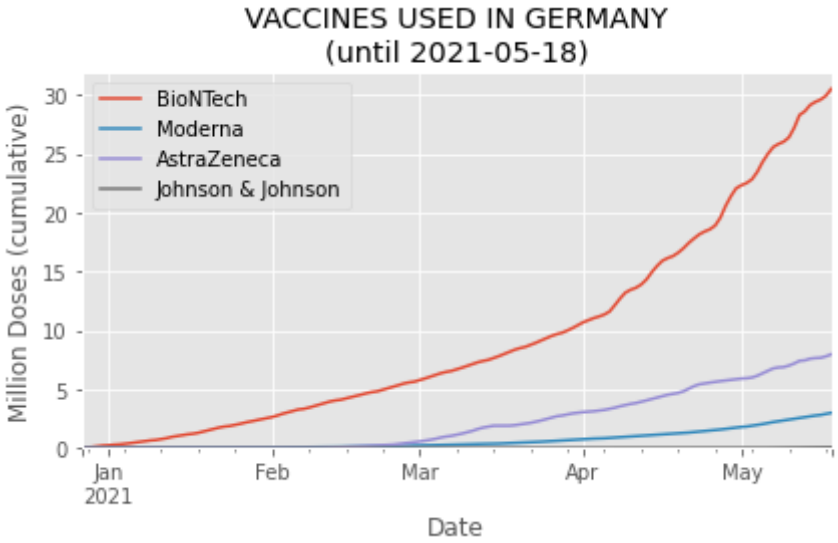
```
vaccine_use = vaccinations.loc[ : , ['date', 'dosen_biontech_kumulativ',
                                     'dosen_moderna_kumulativ',
                                     'dosen_astrazeneca_kumulativ',
                                     'dosen_johnson_kumulativ']]
# Rename columns
vaccine_use.columns = ['date', 'BioNTech', 'Moderna', 'AstraZeneca', 'Johnson & Johnson']
# make 'date' an index
vaccine_use.set_index('date', inplace=True)
# divide columns by 1 million
vaccine_use["BioNTech"] = vaccine_use["BioNTech"] / 1_000_000
vaccine_use["Moderna"] = vaccine_use["Moderna"] / 1_000_000
vaccine_use["AstraZeneca"] = vaccine_use["AstraZeneca"] / 1_000_000
vaccine_use["Johnson & Johnson"] = vaccine_use["Johnson & Johnson"] / 1_000_000
vaccine_use.tail(3)
```

Out[114…

date	BioNTech	Moderna	AstraZeneca	Johnson & Johnson
2021-05-16	29.609383	2.812217	7.669820	0.049847
2021-05-17	29.936142	2.892237	7.788613	0.056879
2021-05-18	30.511166	2.986323	7.957797	0.062563

In [115…

```
vaccines_used = vaccine_use.plot(
    # as it is cumulative, the last row must contain the single highest number
    ylim=(0,math.ceil(max(vaccine_use.iloc[-1]))+1),
    xlabel='Date',
    ylabel='Million Doses (cumulative)',
    title=f"VACCINES USED IN GERMANY\n(until {last_update})")
```



In [116…

```
fig = vaccines_used.get_figure()
fig.savefig('img/vaccines_used_in_germany.png')
```

Vaccination Centers versus Doctor’s Practices

In [117…

```
by_place = vaccinations.loc[ : , ['date', 'dosen_dim_kumulativ', 'dosen_kbv_kumulativ']]
```

```
by_place.columns = ['date', 'vaccination centers', 'practices']
```

In [118…
```
by_place['vaccination centers daily'] = by_place['vaccination centers'].diff()
by_place['practices daily'] = by_place['practices'].diff()
```

In [119…
```
by_place['percentage practices'] =  round(
    by_place['practices daily'] * 100 /
    (by_place['vaccination centers daily'] + by_place['practices daily']), 2)

by_place['percentage centers'] = 100 - by_place['percentage practices']
```

In [120…
```
# make 'date' an index
by_place.set_index('date', inplace=True)
```

In [121…
```
by_place
```

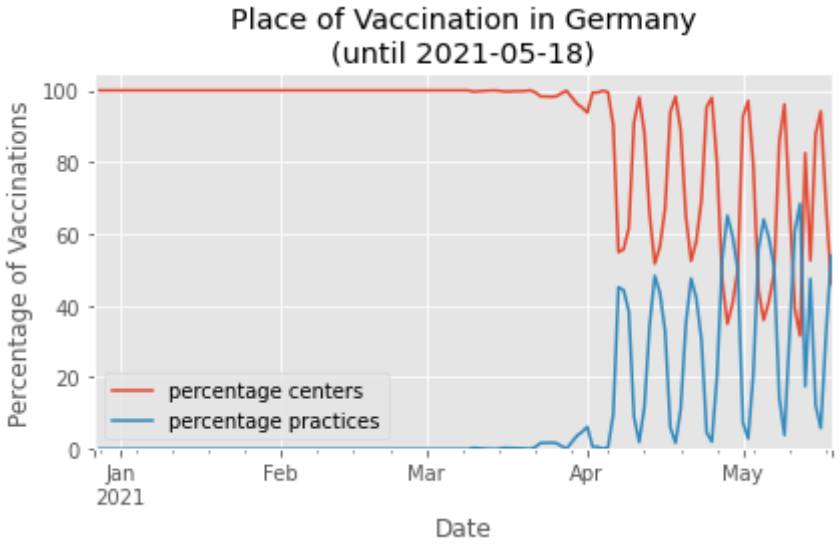
Out[121…

date	vaccination centers	practices	vaccination centers daily	practices daily	percentage practices	percentage centers
2020-12-27	24087	0	NaN	NaN	NaN	NaN
2020-12-28	42647	0	18560.0	0.0	0.00	100.00
2020-12-29	93471	0	50824.0	0.0	0.00	100.00
2020-12-30	156253	0	62782.0	0.0	0.00	100.00
2020-12-31	206444	0	50191.0	0.0	0.00	100.00
...	...	...	...	...	...	...
2021-05-14	29449478	10028800	388889.0	351337.0	47.46	52.54
2021-05-15	29790142	10076466	340664.0	47666.0	12.27	87.73
2021-05-16	30048796	10092471	258654.0	16005.0	5.83	94.17
2021-05-17	30412082	10261789	363286.0	169318.0	31.79	68.21
2021-05-18	30801056	10716793	388974.0	455004.0	53.91	46.09

143 rows × 6 columns

In [122…
```
share = by_place.loc[ : , ['percentage centers', 'percentage practices']]
```

In [123…
```
vacc_shares = share.plot(
    # as it is cumulative, the last row must contain the single highest number
    ylim=(0, 105),  # above 100 to see the line
    xlabel='Date',
    ylabel='Percentage of Vaccinations',
    title=f"Place of Vaccination in Germany\n(until {last_update})")
```



In [124…
```
fig = vacc_shares.get_figure()
fig.savefig('img/vaccinations_germany_by_place.png')
```

Other units of Time

In [125…
```
by_place_daily = by_place.loc[ : , ['vaccination centers daily', 'practices daily']]
by_place_daily.columns = ['vaccination centers', 'practices']
by_place_daily.reset_index(inplace=True)
```

Monthly

In [126…
```
by_place_monthly = by_place_daily.groupby(pd.Grouper(key='date',freq='M')).sum()
by_place_monthly.tail()
```

Out[126…

date	vaccination centers	practices

date	vaccination centers	practices
2021-01-31	2344521.0	0.0
2021-02-28	3778865.0	0.0
2021-03-31	7786287.0	66234.0
2021-04-30	10182842.0	5329140.0
2021-05-31	6502097.0	5321419.0

Scale:

In [127…
```
by_place_monthly['vaccination centers'] = by_place_monthly['vaccination centers'] / 1_000_000
by_place_monthly['practices'] = by_place_monthly['practices'] / 1_000_000
```

Rename the columns

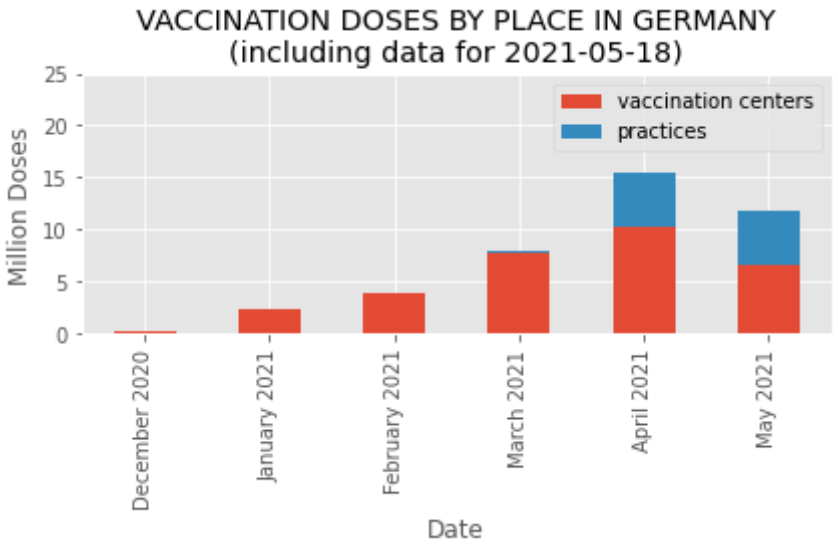
In [128…
```
by_place_monthly['month'] = by_place_monthly.index.strftime('%B')
by_place_monthly['year'] =  by_place_monthly.index.strftime('%Y')
by_place_monthly['label'] = by_place_monthly['month'] + ' ' + by_place_monthly['year']
by_place_monthly.drop(columns=['month', 'year'], inplace=True)
by_place_monthly.set_index('label', inplace=True)
by_place_monthly.tail(6)
```

Out[128…

label	vaccination centers	practices
December 2020	0.182357	0.000000
January 2021	2.344521	0.000000
February 2021	3.778865	0.000000
March 2021	7.786287	0.066234
April 2021	10.182842	5.329140
May 2021	6.502097	5.321419

In [129…
```
monthly_plot = by_place_monthly.plot.bar(
    stacked=True,
```

```
    ylim=(0, 25),
    xlabel='Date',
    ylabel='Million Doses',
    title=f"VACCINATION DOSES BY PLACE IN GERMANY\n(including data for {last_update})")
```



In [130…

```
fig = monthly_plot.get_figure()
fig.savefig('img/monthly_doses_by_place_germany.png')
```