

lab2

May 18, 2023

1 - Załadowanie biblioteki matplotlib i innych potrzebnych bibliotek

```
[4]: %matplotlib inline

import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
```

2 - załadowanie danych

```
[5]: df = pd.read_csv("IHME_GBD_2019_SMOKING_TOB_1990_2019_NUM_SMOKERS_Y2021M05D27.
↳csv", encoding = "utf-8")
df
```

```
[5]:
```

	measure_name	location_id	location_name	sex_id	sex_name	\
0	Number of Smokers	1	Global	1	Male	
1	Number of Smokers	1	Global	2	Female	
2	Number of Smokers	1	Global	3	Both	
3	Number of Smokers	1	Global	1	Male	
4	Number of Smokers	1	Global	2	Female	
...	
20965	Number of Smokers	522	Sudan	2	Female	
20966	Number of Smokers	522	Sudan	3	Both	
20967	Number of Smokers	522	Sudan	1	Male	
20968	Number of Smokers	522	Sudan	2	Female	
20969	Number of Smokers	522	Sudan	3	Both	

	age_group_id	age_group_name	year_id	val	upper	\
0	29	15+ years	1990	8.031015e+08	8.096221e+08	
1	29	15+ years	1990	1.891488e+08	1.930929e+08	
2	29	15+ years	1990	9.922503e+08	1.000161e+09	
3	29	15+ years	1991	8.138972e+08	8.200339e+08	
4	29	15+ years	1991	1.905375e+08	1.944249e+08	
...	
20965	29	15+ years	2018	2.435999e+05	3.286166e+05	
20966	29	15+ years	2018	2.610672e+06	2.833943e+06	
20967	29	15+ years	2019	2.439150e+06	2.656579e+06	
20968	29	15+ years	2019	2.500800e+05	3.345384e+05	

```
20969          29      15+ years      2019  2.689230e+06  2.918332e+06
```

```

              lower
0      7.959086e+08
1      1.855595e+08
2      9.847880e+08
3      8.069514e+08
4      1.869744e+08
...
20965  1.752508e+05
20966  2.409108e+06
20967  2.236450e+06
20968  1.816686e+05
20969  2.480656e+06
```

```
[20970 rows x 11 columns]
```

3 - użycie matplotlib

```
[3]: matplotlib = df[(df["sex_name"] == "Both") & (df["location_name"] == "Poland")]
matplotlib
```

```
[3]:
      measure_name  location_id location_name  sex_id sex_name \
4322 Number of Smokers           51      Poland         3    Both
4325 Number of Smokers           51      Poland         3    Both
4328 Number of Smokers           51      Poland         3    Both
4331 Number of Smokers           51      Poland         3    Both
4334 Number of Smokers           51      Poland         3    Both
4337 Number of Smokers           51      Poland         3    Both
4340 Number of Smokers           51      Poland         3    Both
4343 Number of Smokers           51      Poland         3    Both
4346 Number of Smokers           51      Poland         3    Both
4349 Number of Smokers           51      Poland         3    Both
4352 Number of Smokers           51      Poland         3    Both
4355 Number of Smokers           51      Poland         3    Both
4358 Number of Smokers           51      Poland         3    Both
4361 Number of Smokers           51      Poland         3    Both
4364 Number of Smokers           51      Poland         3    Both
4367 Number of Smokers           51      Poland         3    Both
4370 Number of Smokers           51      Poland         3    Both
4373 Number of Smokers           51      Poland         3    Both
4376 Number of Smokers           51      Poland         3    Both
4379 Number of Smokers           51      Poland         3    Both
4382 Number of Smokers           51      Poland         3    Both
4385 Number of Smokers           51      Poland         3    Both
4388 Number of Smokers           51      Poland         3    Both
4391 Number of Smokers           51      Poland         3    Both
```

4394	Number of Smokers	51	Poland	3	Both
4397	Number of Smokers	51	Poland	3	Both
4400	Number of Smokers	51	Poland	3	Both
4403	Number of Smokers	51	Poland	3	Both
4406	Number of Smokers	51	Poland	3	Both
4409	Number of Smokers	51	Poland	3	Both

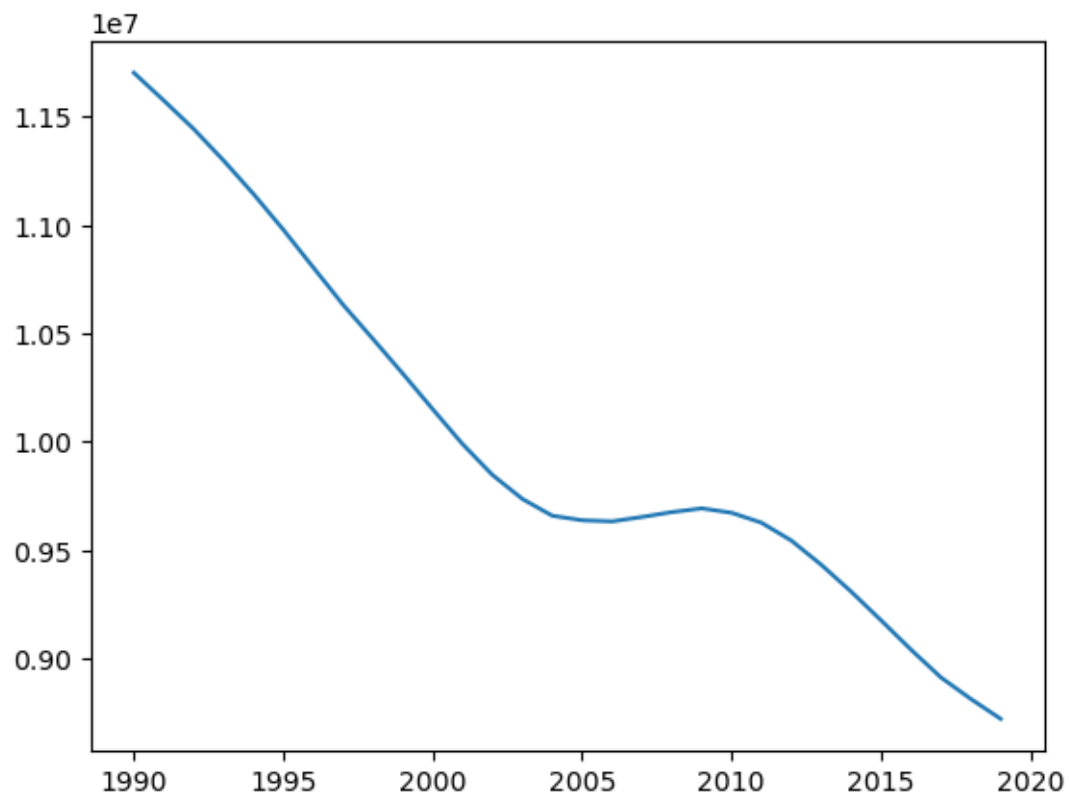
	age_group_id	age_group_name	year_id	val	upper \
4322	29	15+ years	1990	1.170448e+07	1.201222e+07
4325	29	15+ years	1991	1.157647e+07	1.188570e+07
4328	29	15+ years	1992	1.144506e+07	1.175827e+07
4331	29	15+ years	1993	1.129936e+07	1.162928e+07
4334	29	15+ years	1994	1.114488e+07	1.147066e+07
4337	29	15+ years	1995	1.097960e+07	1.132458e+07
4340	29	15+ years	1996	1.080558e+07	1.115800e+07
4343	29	15+ years	1997	1.063264e+07	1.098329e+07
4346	29	15+ years	1998	1.047442e+07	1.083031e+07
4349	29	15+ years	1999	1.031418e+07	1.068428e+07
4352	29	15+ years	2000	1.014975e+07	1.051794e+07
4355	29	15+ years	2001	9.988362e+06	1.032469e+07
4358	29	15+ years	2002	9.845932e+06	1.014307e+07
4361	29	15+ years	2003	9.735058e+06	1.002336e+07
4364	29	15+ years	2004	9.657597e+06	9.922015e+06
4367	29	15+ years	2005	9.636582e+06	9.878019e+06
4370	29	15+ years	2006	9.631582e+06	9.875923e+06
4373	29	15+ years	2007	9.651840e+06	9.889532e+06
4376	29	15+ years	2008	9.674385e+06	9.914297e+06
4379	29	15+ years	2009	9.691581e+06	9.917740e+06
4382	29	15+ years	2010	9.671307e+06	9.904838e+06
4385	29	15+ years	2011	9.625299e+06	9.869306e+06
4388	29	15+ years	2012	9.542767e+06	9.789413e+06
4391	29	15+ years	2013	9.430694e+06	9.689100e+06
4394	29	15+ years	2014	9.306823e+06	9.564203e+06
4397	29	15+ years	2015	9.174107e+06	9.438002e+06
4400	29	15+ years	2016	9.038816e+06	9.326490e+06
4403	29	15+ years	2017	8.910715e+06	9.215457e+06
4406	29	15+ years	2018	8.811197e+06	9.135481e+06
4409	29	15+ years	2019	8.719168e+06	9.075511e+06

	lower
4322	1.138875e+07
4325	1.125896e+07
4328	1.112616e+07
4331	1.097590e+07
4334	1.080890e+07
4337	1.062839e+07
4340	1.045887e+07

```
4343 1.027121e+07
4346 1.012083e+07
4349 9.960286e+06
4352 9.794037e+06
4355 9.655642e+06
4358 9.532336e+06
4361 9.443244e+06
4364 9.384474e+06
4367 9.380627e+06
4370 9.394114e+06
4373 9.426812e+06
4376 9.444482e+06
4379 9.441793e+06
4382 9.416823e+06
4385 9.364238e+06
4388 9.278873e+06
4391 9.175417e+06
4394 9.048851e+06
4397 8.910913e+06
4400 8.764274e+06
4403 8.620903e+06
4406 8.490369e+06
4409 8.370563e+06
```

```
[4]: plt.plot(matplot["year_id"], matplot["val"])
```

```
[4]: [<matplotlib.lines.Line2D at 0x261892ac340>]
```



4 - style rysowanie

```
[5]: plt.plot(matplot["year_id"], matplot["val"], 'k--')
```

```
[5]: [<matplotlib.lines.Line2D at 0x2618938d600>]
```



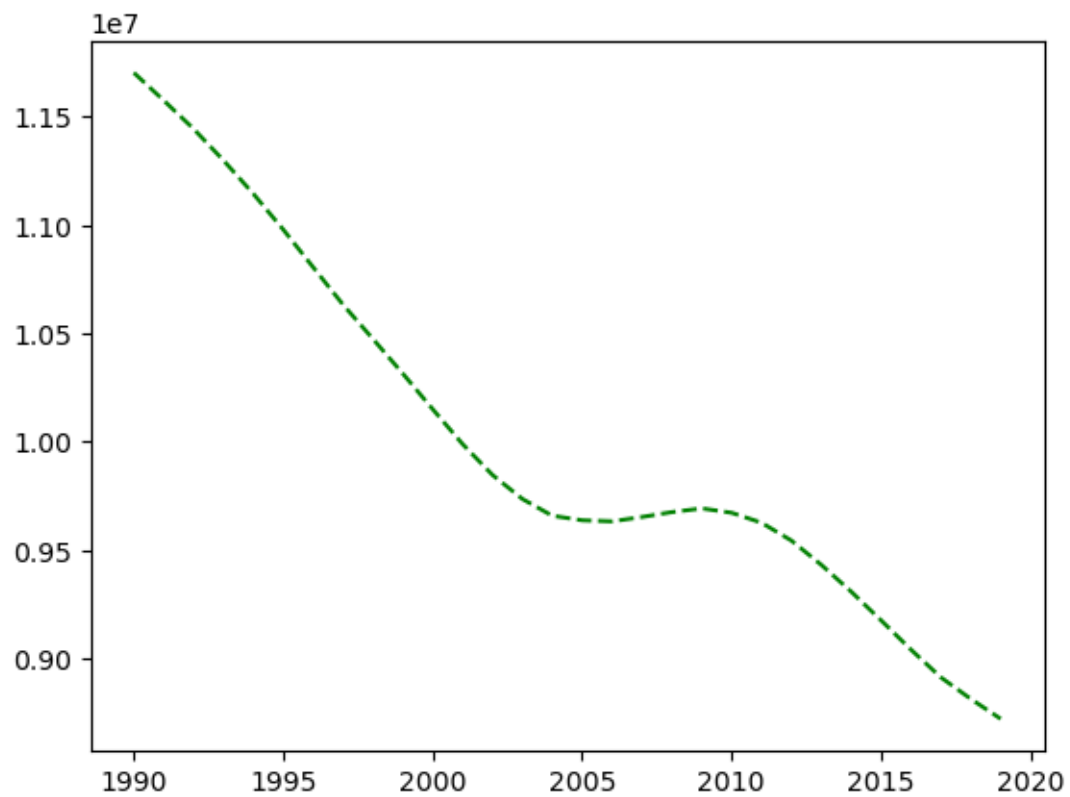
```
[6]: plt.plot(matplot["year_id"], matplot["val"], 'k-')
```

```
[6]: [<matplotlib.lines.Line2D at 0x26189bd33d0>]
```



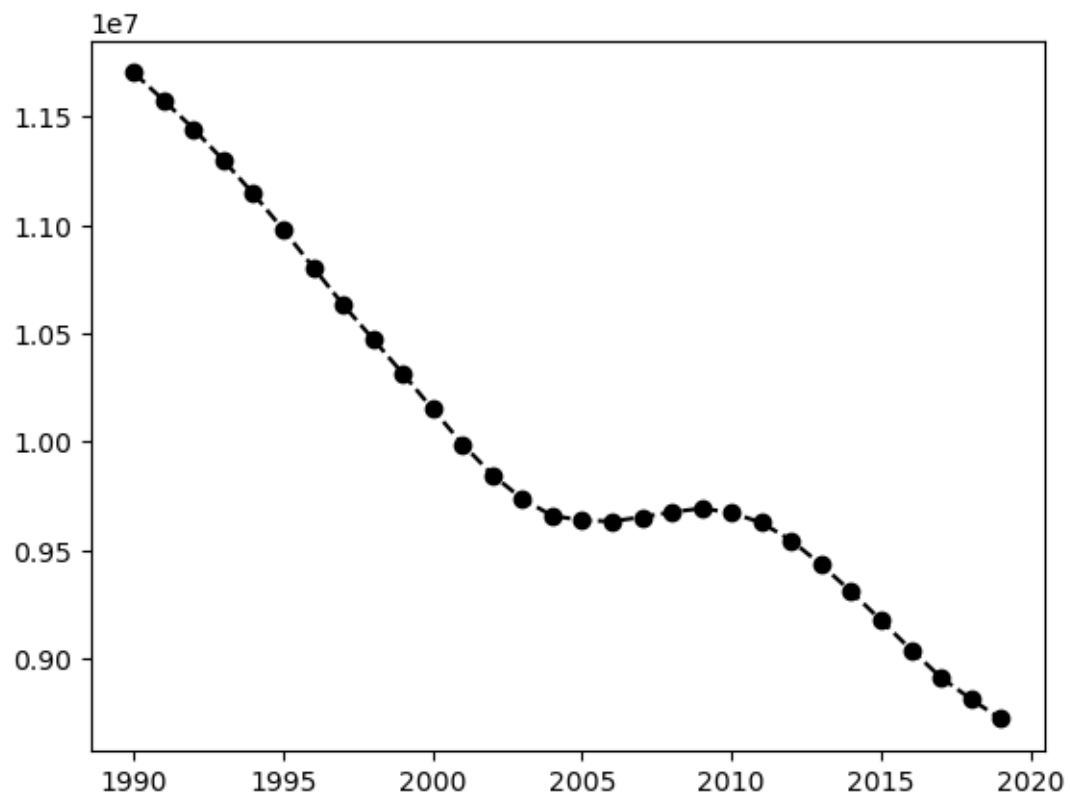
```
[10]: plt.plot(matplot["year_id"], matplot["val"], 'g--')
```

```
[10]: [<matplotlib.lines.Line2D at 0x26189479870>]
```



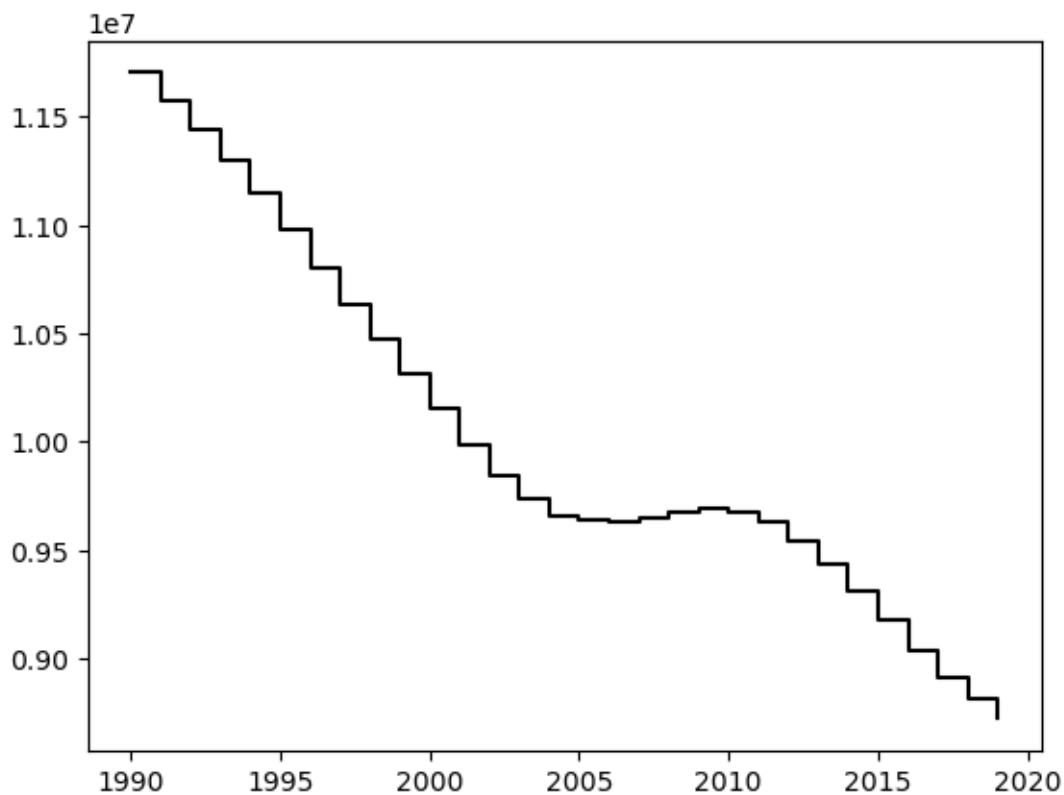
```
[11]: plt.plot(matplot["year_id"], matplot["val"], 'ko--')
```

```
[11]: [<matplotlib.lines.Line2D at 0x26189ea7d60>]
```

```
[12]: plt.plot(matplot["year_id"], matplot["val"], 'k-', drawstyle='steps-post',
↳ label='steps-post')
```

```
[12]: [<matplotlib.lines.Line2D at 0x2618a7fd870>]
```



5 - zapisywanie rysunków do pliku

```
[13]: plt.savefig('figpath.png', dpi=400, bbox_inches='tight')
```

<Figure size 640x480 with 0 Axes>

6 - wykresy liniowe

```
[14]: liniowy_series = df[(df["sex_name"] == "Both") & (df["location_name"] == "Poland")]
liniowy_series
```

```
[14]:
```

	measure_name	location_id	location_name	sex_id	sex_name	\
4322	Number of Smokers	51	Poland	3	Both	
4325	Number of Smokers	51	Poland	3	Both	
4328	Number of Smokers	51	Poland	3	Both	
4331	Number of Smokers	51	Poland	3	Both	
4334	Number of Smokers	51	Poland	3	Both	
4337	Number of Smokers	51	Poland	3	Both	
4340	Number of Smokers	51	Poland	3	Both	
4343	Number of Smokers	51	Poland	3	Both	
4346	Number of Smokers	51	Poland	3	Both	

4349	Number of Smokers	51	Poland	3	Both
4352	Number of Smokers	51	Poland	3	Both
4355	Number of Smokers	51	Poland	3	Both
4358	Number of Smokers	51	Poland	3	Both
4361	Number of Smokers	51	Poland	3	Both
4364	Number of Smokers	51	Poland	3	Both
4367	Number of Smokers	51	Poland	3	Both
4370	Number of Smokers	51	Poland	3	Both
4373	Number of Smokers	51	Poland	3	Both
4376	Number of Smokers	51	Poland	3	Both
4379	Number of Smokers	51	Poland	3	Both
4382	Number of Smokers	51	Poland	3	Both
4385	Number of Smokers	51	Poland	3	Both
4388	Number of Smokers	51	Poland	3	Both
4391	Number of Smokers	51	Poland	3	Both
4394	Number of Smokers	51	Poland	3	Both
4397	Number of Smokers	51	Poland	3	Both
4400	Number of Smokers	51	Poland	3	Both
4403	Number of Smokers	51	Poland	3	Both
4406	Number of Smokers	51	Poland	3	Both
4409	Number of Smokers	51	Poland	3	Both

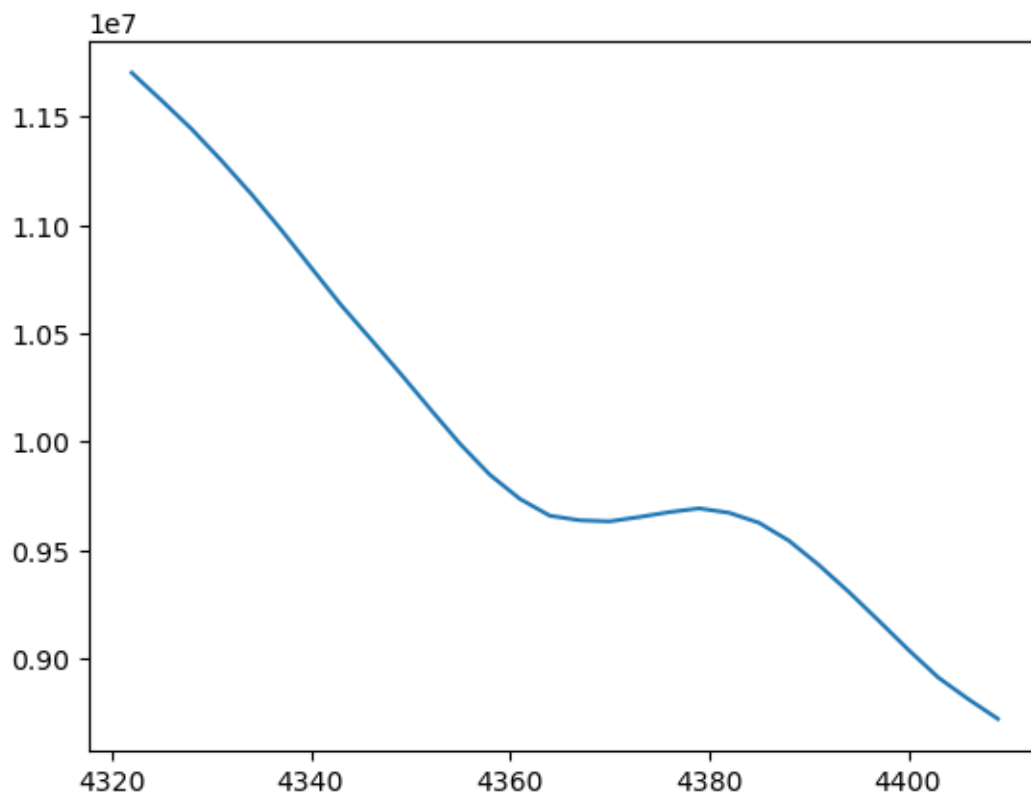
	age_group_id	age_group_name	year_id	val	upper \
4322	29	15+ years	1990	1.170448e+07	1.201222e+07
4325	29	15+ years	1991	1.157647e+07	1.188570e+07
4328	29	15+ years	1992	1.144506e+07	1.175827e+07
4331	29	15+ years	1993	1.129936e+07	1.162928e+07
4334	29	15+ years	1994	1.114488e+07	1.147066e+07
4337	29	15+ years	1995	1.097960e+07	1.132458e+07
4340	29	15+ years	1996	1.080558e+07	1.115800e+07
4343	29	15+ years	1997	1.063264e+07	1.098329e+07
4346	29	15+ years	1998	1.047442e+07	1.083031e+07
4349	29	15+ years	1999	1.031418e+07	1.068428e+07
4352	29	15+ years	2000	1.014975e+07	1.051794e+07
4355	29	15+ years	2001	9.988362e+06	1.032469e+07
4358	29	15+ years	2002	9.845932e+06	1.014307e+07
4361	29	15+ years	2003	9.735058e+06	1.002336e+07
4364	29	15+ years	2004	9.657597e+06	9.922015e+06
4367	29	15+ years	2005	9.636582e+06	9.878019e+06
4370	29	15+ years	2006	9.631582e+06	9.875923e+06
4373	29	15+ years	2007	9.651840e+06	9.889532e+06
4376	29	15+ years	2008	9.674385e+06	9.914297e+06
4379	29	15+ years	2009	9.691581e+06	9.917740e+06
4382	29	15+ years	2010	9.671307e+06	9.904838e+06
4385	29	15+ years	2011	9.625299e+06	9.869306e+06
4388	29	15+ years	2012	9.542767e+06	9.789413e+06
4391	29	15+ years	2013	9.430694e+06	9.689100e+06

4394	29	15+ years	2014	9.306823e+06	9.564203e+06
4397	29	15+ years	2015	9.174107e+06	9.438002e+06
4400	29	15+ years	2016	9.038816e+06	9.326490e+06
4403	29	15+ years	2017	8.910715e+06	9.215457e+06
4406	29	15+ years	2018	8.811197e+06	9.135481e+06
4409	29	15+ years	2019	8.719168e+06	9.075511e+06

	lower
4322	1.138875e+07
4325	1.125896e+07
4328	1.112616e+07
4331	1.097590e+07
4334	1.080890e+07
4337	1.062839e+07
4340	1.045887e+07
4343	1.027121e+07
4346	1.012083e+07
4349	9.960286e+06
4352	9.794037e+06
4355	9.655642e+06
4358	9.532336e+06
4361	9.443244e+06
4364	9.384474e+06
4367	9.380627e+06
4370	9.394114e+06
4373	9.426812e+06
4376	9.444482e+06
4379	9.441793e+06
4382	9.416823e+06
4385	9.364238e+06
4388	9.278873e+06
4391	9.175417e+06
4394	9.048851e+06
4397	8.910913e+06
4400	8.764274e+06
4403	8.620903e+06
4406	8.490369e+06
4409	8.370563e+06

```
[15]: w liniowy_series = pd.Series(liniowy_series["val"])
      w liniowy_series.plot()
```

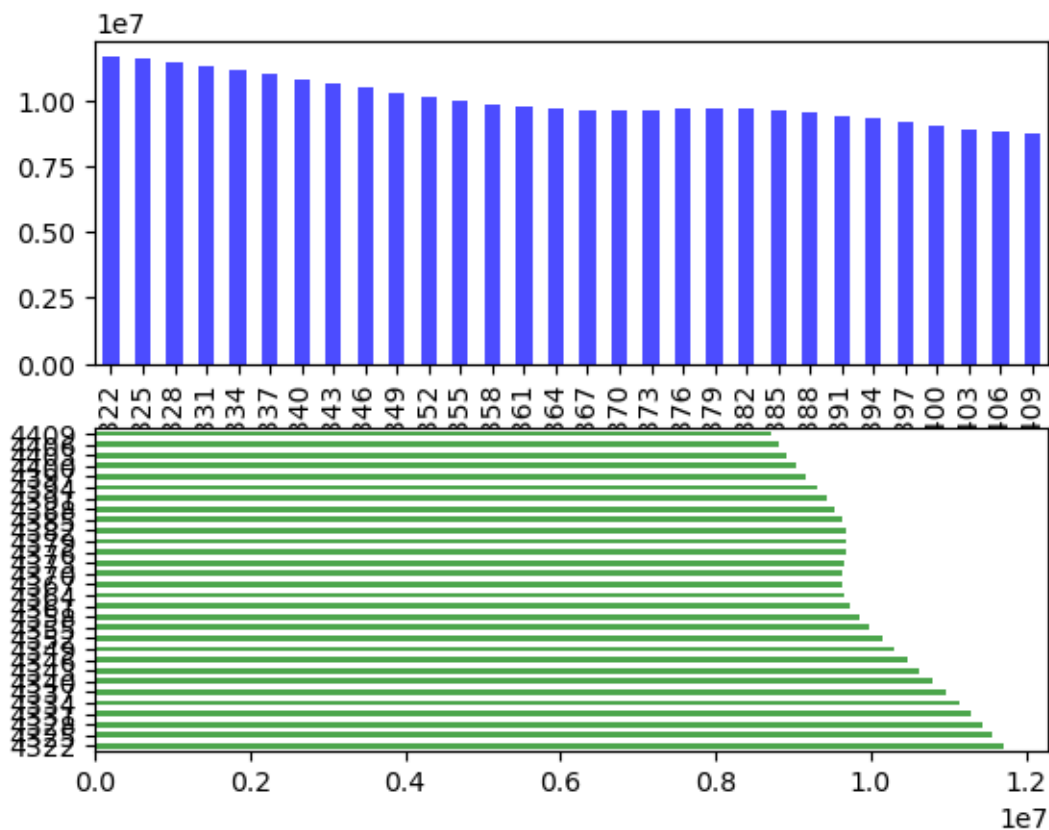
```
[15]: <Axes: >
```



7 - wykresy kolumnowe

```
[19]: fig, axes = plt.subplots(2, 1)
data = df[(df["sex_name"] == "Both") & (df["location_name"] == "Poland")]
data = pd.Series(liniowy_series["val"])
data.plot.bar(ax=axes[0], color='b', alpha=0.7)
data.plot.barh(ax=axes[1], color='g', alpha=0.7)
```

[19]: <Axes: >



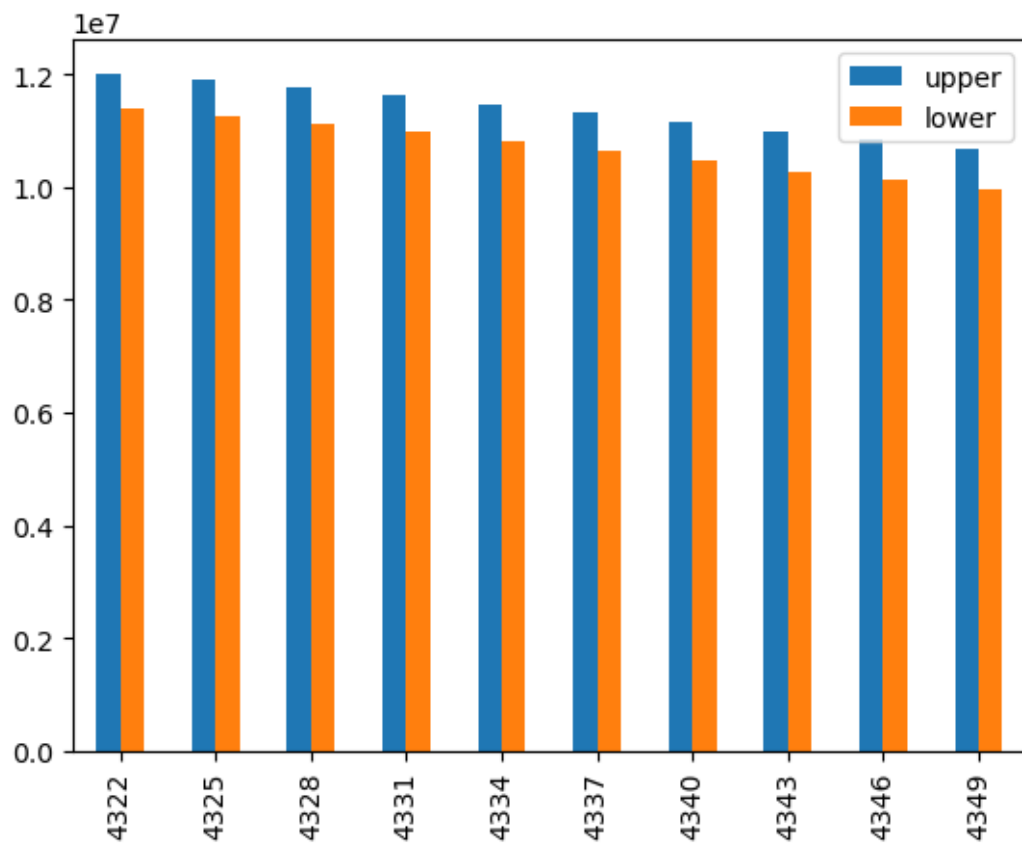
```
[23]: w_kolumnowy = df[(df["sex_name"] == "Both") & (df["location_name"] == "Poland")]
w_kolumnowy = w_kolumnowy[["upper", "lower"]]
w_kolumnowy = w_kolumnowy.head(10)
w_kolumnowy
```

```
[23]:
```

	upper	lower
4322	12012219.84	1.138875e+07
4325	11885703.34	1.125896e+07
4328	11758271.45	1.112616e+07
4331	11629281.89	1.097590e+07
4334	11470659.42	1.080890e+07
4337	11324575.92	1.062839e+07
4340	11158004.92	1.045887e+07
4343	10983286.74	1.027121e+07
4346	10830307.69	1.012083e+07
4349	10684277.63	9.960286e+06

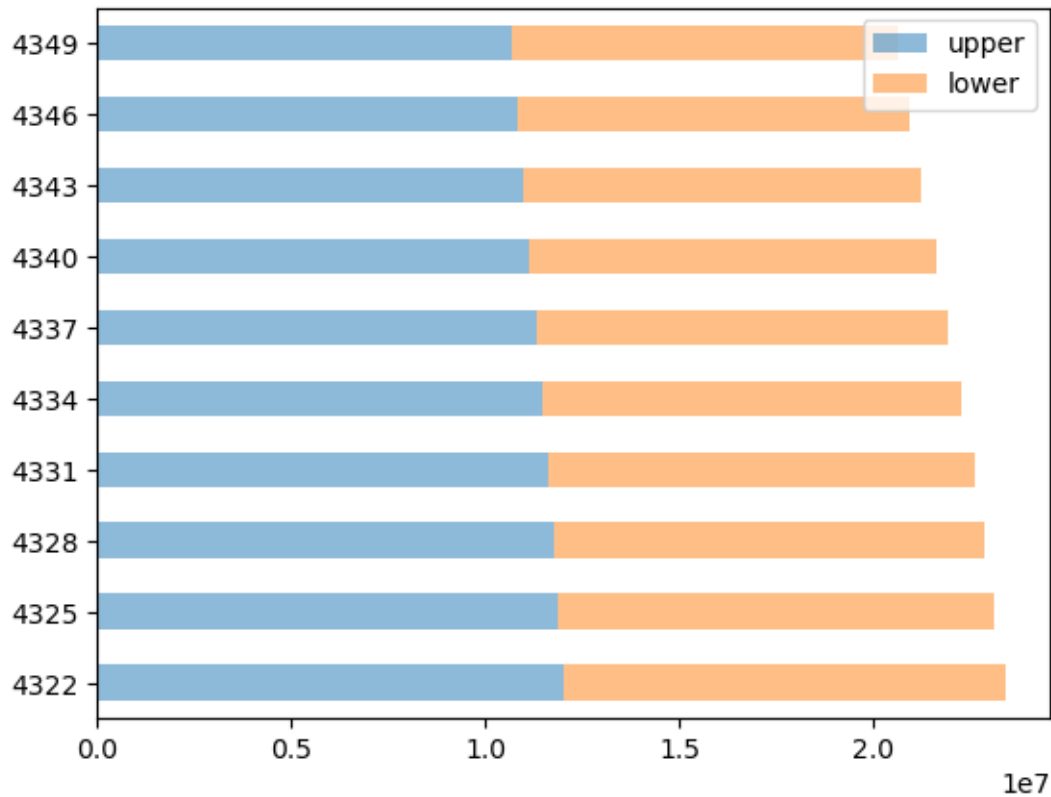
```
[24]: w_kolumnowy.plot.bar()
```

```
[24]: <Axes: >
```



```
[26]: w_kolumnowy.plot.barh(stacked=True, alpha=0.5)
```

```
[26]: <Axes: >
```



8 - seaborn

```
[10]: import seaborn as sns
```

```
[8]: sb = df
sb = sb[sb["year_id"] > 2000]
sb
```

```
[8]:
```

	measure_name	location_id	location_name	sex_id	sex_name	\
33	Number of Smokers	1	Global	1	Male	
34	Number of Smokers	1	Global	2	Female	
35	Number of Smokers	1	Global	3	Both	
36	Number of Smokers	1	Global	1	Male	
37	Number of Smokers	1	Global	2	Female	
...	
20965	Number of Smokers	522	Sudan	2	Female	
20966	Number of Smokers	522	Sudan	3	Both	
20967	Number of Smokers	522	Sudan	1	Male	
20968	Number of Smokers	522	Sudan	2	Female	
20969	Number of Smokers	522	Sudan	3	Both	

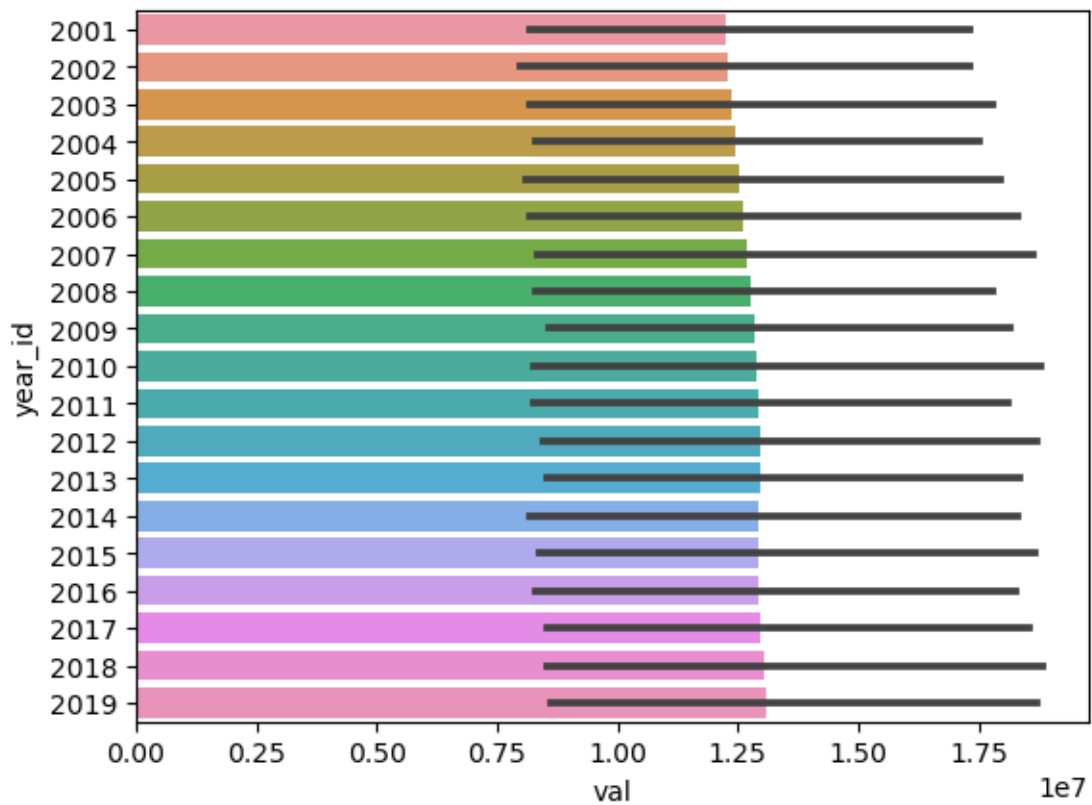
	age_group_id	age_group_name	year_id	val	upper	\
33	29	15+ years	2001	8.653357e+08	8.709454e+08	
34	29	15+ years	2001	2.043228e+08	2.078411e+08	
35	29	15+ years	2001	1.069659e+09	1.076377e+09	
36	29	15+ years	2002	8.697218e+08	8.753433e+08	
37	29	15+ years	2002	2.051251e+08	2.086136e+08	
...	
20965	29	15+ years	2018	2.435999e+05	3.286166e+05	
20966	29	15+ years	2018	2.610672e+06	2.833943e+06	
20967	29	15+ years	2019	2.439150e+06	2.656579e+06	
20968	29	15+ years	2019	2.500800e+05	3.345384e+05	
20969	29	15+ years	2019	2.689230e+06	2.918332e+06	

	lower
33	8.597397e+08
34	2.011226e+08
35	1.062829e+09
36	8.642284e+08
37	2.019433e+08
...	...
20965	1.752508e+05
20966	2.409108e+06
20967	2.236450e+06
20968	1.816686e+05
20969	2.480656e+06

[13281 rows x 11 columns]

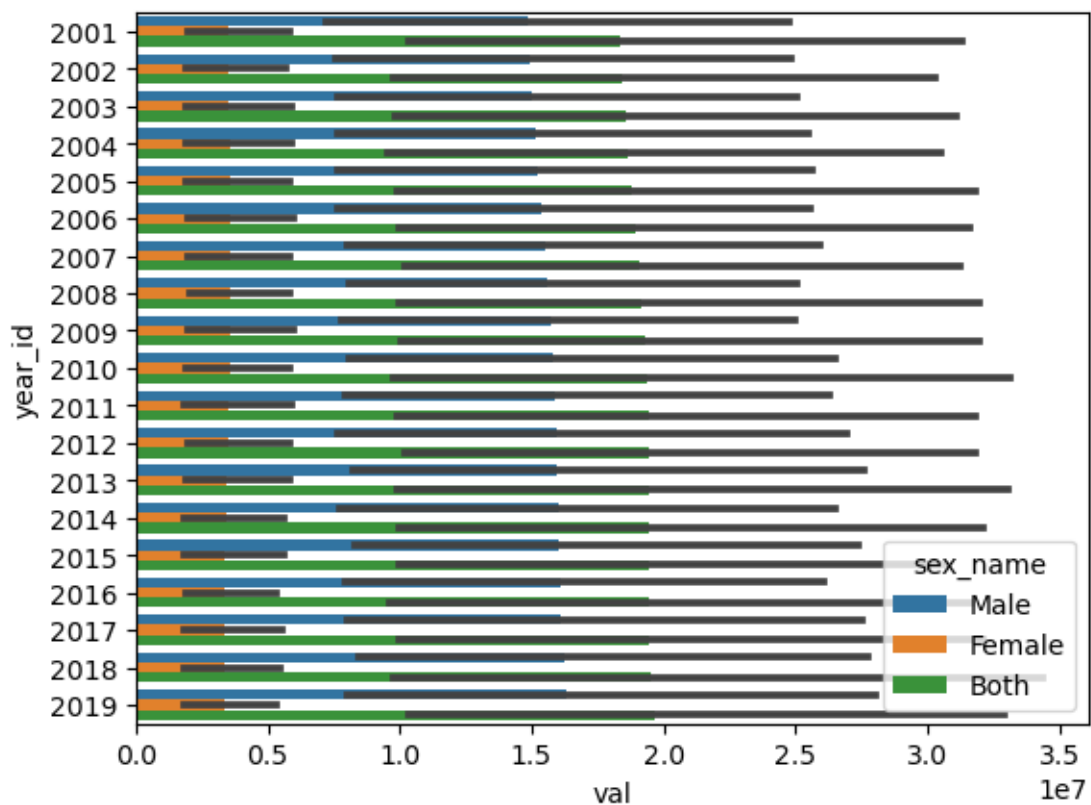
```
[11]: sns.barplot(x='val', y='year_id', data=sb, orient='h')
```

```
[11]: <Axes: xlabel='val', ylabel='year_id'>
```



```
[12]: sns.barplot(x='val', y='year_id', hue= "sex_name", data=sb, orient='h')
```

```
[12]: <Axes: xlabel='val', ylabel='year_id'>
```



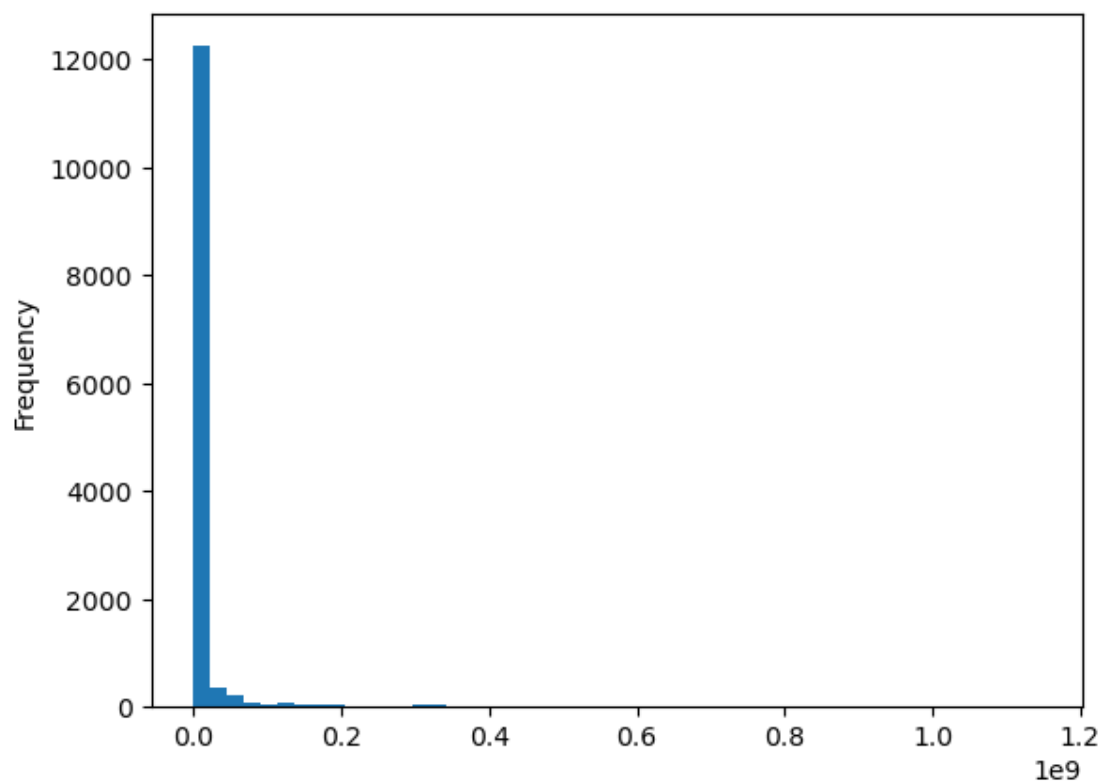
9 - histogramy i wykresy gęstości rozkładu

```
[21]: histogram = df[df["year_id"] > 2000]
      histogram = histogram["val"]
      histogram
```

```
[21]: 33      8.653357e+08
      34      2.043228e+08
      35      1.069659e+09
      36      8.697218e+08
      37      2.051251e+08
      ...
      20965   2.435999e+05
      20966   2.610672e+06
      20967   2.439150e+06
      20968   2.500800e+05
      20969   2.689230e+06
      Name: val, Length: 13281, dtype: float64
```

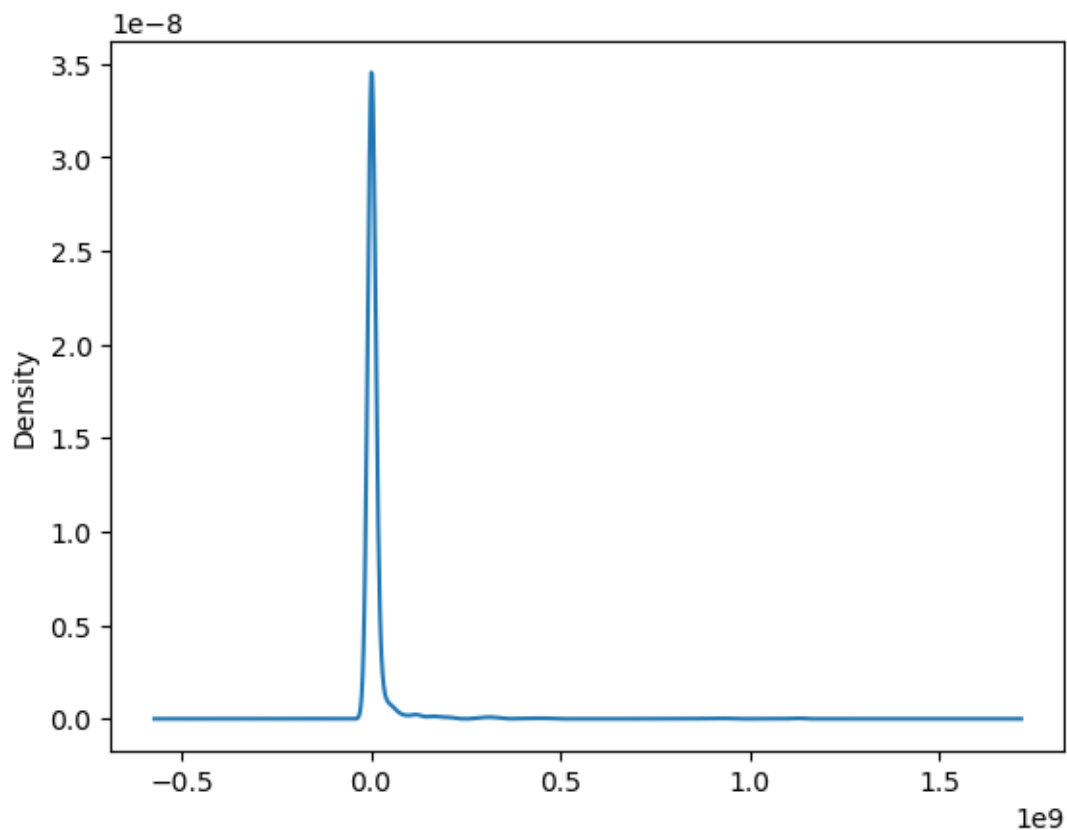
```
[22]: histogram.plot.hist(bins=50)
```

[22]: <Axes: ylabel='Frequency'>



[23]: histogram.plot.kde()

[23]: <Axes: ylabel='Density'>



10 - wykresy punktowe i wykresy bitowe

```
[24]: punkt = df
      punkt = punkt[punkt["year_id"] > 2000]
      punkt
```

```
[24]:
```

	measure_name	location_id	location_name	sex_id	sex_name	\
33	Number of Smokers	1	Global	1	Male	
34	Number of Smokers	1	Global	2	Female	
35	Number of Smokers	1	Global	3	Both	
36	Number of Smokers	1	Global	1	Male	
37	Number of Smokers	1	Global	2	Female	
...	
20965	Number of Smokers	522	Sudan	2	Female	
20966	Number of Smokers	522	Sudan	3	Both	
20967	Number of Smokers	522	Sudan	1	Male	
20968	Number of Smokers	522	Sudan	2	Female	
20969	Number of Smokers	522	Sudan	3	Both	

	age_group_id	age_group_name	year_id	val	upper	\
33	29	15+ years	2001	8.653357e+08	8.709454e+08	

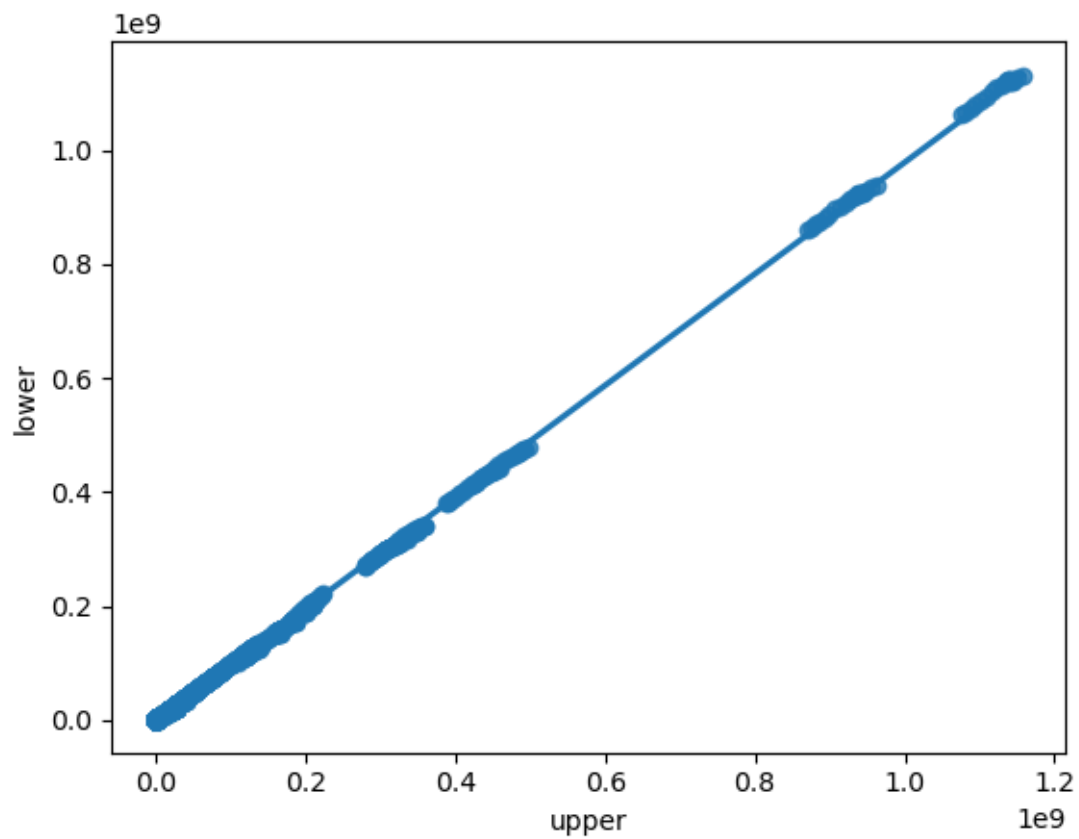
34	29	15+ years	2001	2.043228e+08	2.078411e+08
35	29	15+ years	2001	1.069659e+09	1.076377e+09
36	29	15+ years	2002	8.697218e+08	8.753433e+08
37	29	15+ years	2002	2.051251e+08	2.086136e+08
...
20965	29	15+ years	2018	2.435999e+05	3.286166e+05
20966	29	15+ years	2018	2.610672e+06	2.833943e+06
20967	29	15+ years	2019	2.439150e+06	2.656579e+06
20968	29	15+ years	2019	2.500800e+05	3.345384e+05
20969	29	15+ years	2019	2.689230e+06	2.918332e+06

	lower
33	8.597397e+08
34	2.011226e+08
35	1.062829e+09
36	8.642284e+08
37	2.019433e+08
...	...
20965	1.752508e+05
20966	2.409108e+06
20967	2.236450e+06
20968	1.816686e+05
20969	2.480656e+06

[13281 rows x 11 columns]

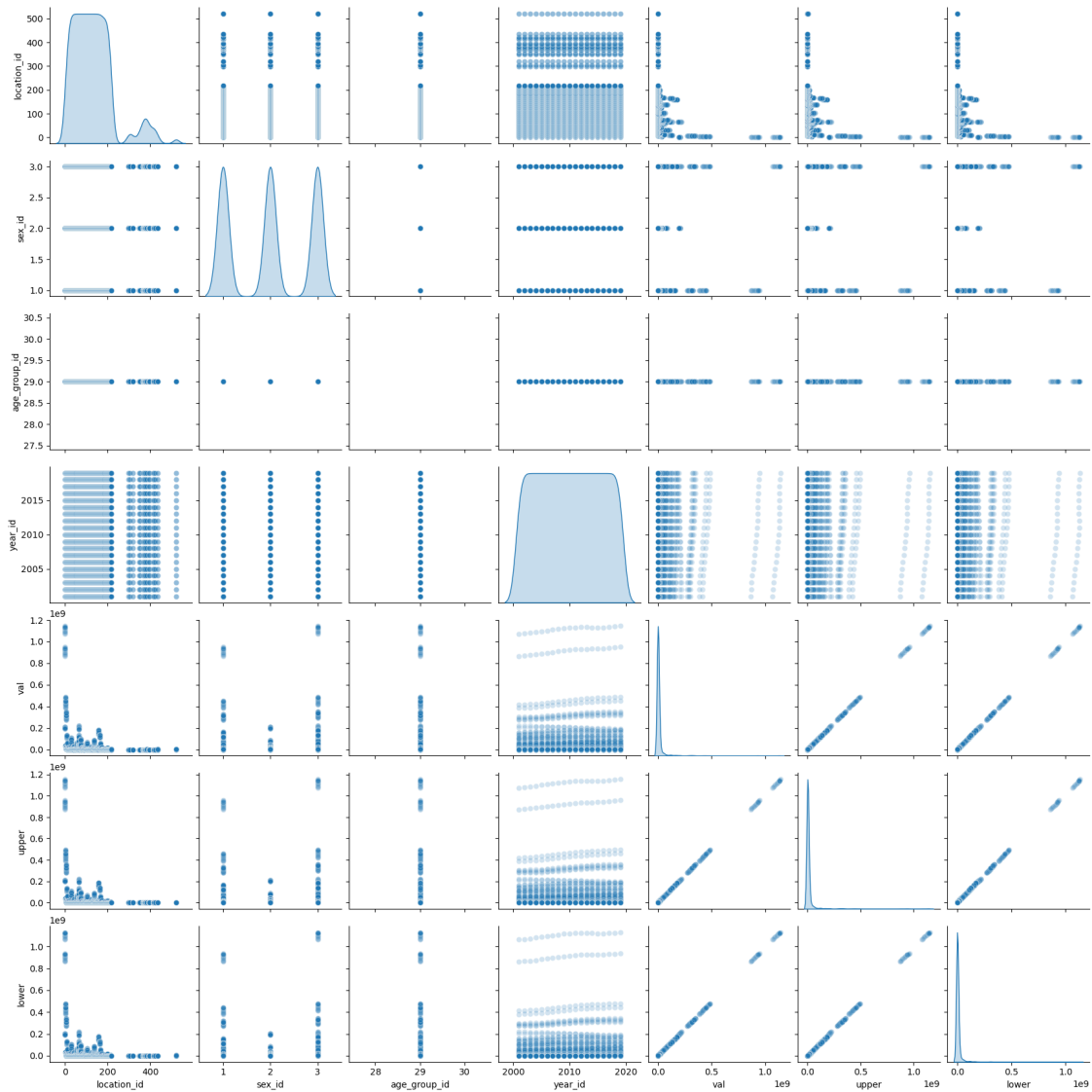
```
[26]: sns.regplot(x='upper', y='lower', data=punkt)
```

```
[26]: <Axes: xlabel='upper', ylabel='lower'>
```



```
[27]: sns.pairplot(punkt, diag_kind='kde', plot_kws={'alpha': 0.2})
```

```
[27]: <seaborn.axisgrid.PairGrid at 0x21902b823b0>
```



11 - dane kategoryczne

```
[32]: katego = df
katego = katego[(katego["location_name"] == "Poland")]
katego
```

```
[32]:
```

	measure_name	location_id	location_name	sex_id	sex_name	\
4320	Number of Smokers	51	Poland	1	Male	
4321	Number of Smokers	51	Poland	2	Female	
4322	Number of Smokers	51	Poland	3	Both	
4323	Number of Smokers	51	Poland	1	Male	
4324	Number of Smokers	51	Poland	2	Female	
...	

4405	Number of Smokers	51	Poland	2	Female
4406	Number of Smokers	51	Poland	3	Both
4407	Number of Smokers	51	Poland	1	Male
4408	Number of Smokers	51	Poland	2	Female
4409	Number of Smokers	51	Poland	3	Both

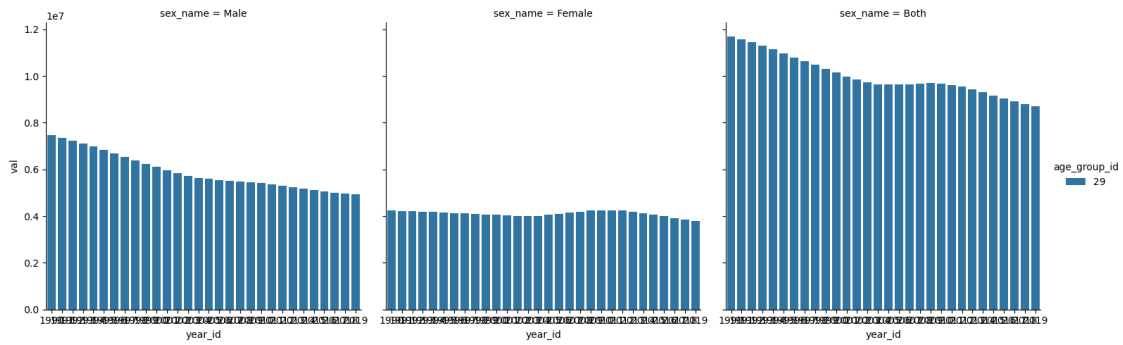
	age_group_id	age_group_name	year_id	val	upper \
4320	29	15+ years	1990	7.468990e+06	7.679963e+06
4321	29	15+ years	1990	4.235491e+06	4.465004e+06
4322	29	15+ years	1990	1.170448e+07	1.201222e+07
4323	29	15+ years	1991	7.358603e+06	7.564003e+06
4324	29	15+ years	1991	4.217866e+06	4.454244e+06
...
4405	29	15+ years	2018	3.848206e+06	4.116964e+06
4406	29	15+ years	2018	8.811197e+06	9.135481e+06
4407	29	15+ years	2019	4.938503e+06	5.166168e+06
4408	29	15+ years	2019	3.780665e+06	4.072287e+06
4409	29	15+ years	2019	8.719168e+06	9.075511e+06

	lower
4320	7.261637e+06
4321	3.987365e+06
4322	1.138875e+07
4323	7.154446e+06
4324	3.968376e+06
...	...
4405	3.595334e+06
4406	8.490369e+06
4407	4.737903e+06
4408	3.504634e+06
4409	8.370563e+06

[90 rows x 11 columns]

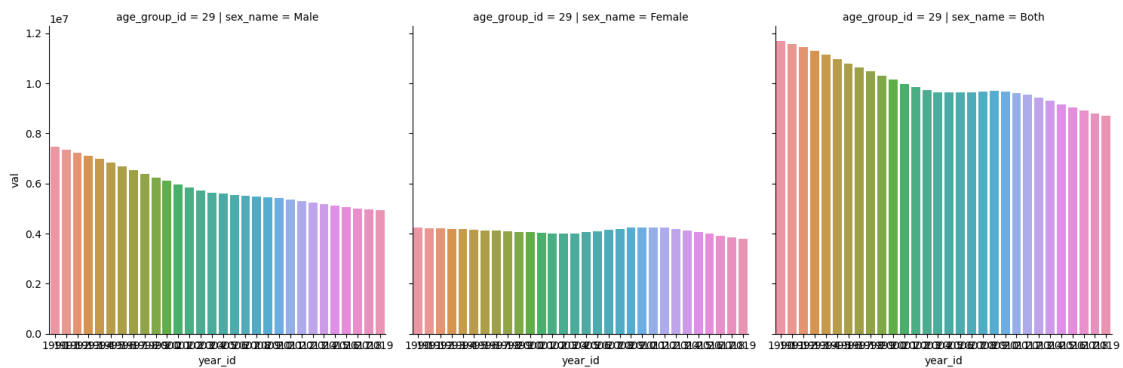
```
[35]: sns.catplot(x='year_id', y='val', hue='age_group_id', col='sex_name',
               kind='bar', data=katego)
```

```
[35]: <seaborn.axisgrid.FacetGrid at 0x21905c0ee60>
```



```
[36]: sns.catplot(x='year_id', y='val', row='age_group_id', col='sex_name',
    ↪ kind='bar', data=katego)
```

```
[36]: <seaborn.axisgrid.FacetGrid at 0x2190adce170>
```



```
[ ]:
```