**Notes on dataset and questions:**

Association between solar radiation at a location and something else.

Pharmaceutical consumption affect by other factor

Relation between use of hypnotics and sedatives and solar radiation

Solar radiation in dataset:

<https://registry.opendata.aws/nasa-power/>

NEW Solar radiation in dataset (includes country codes):

https://datacatalog.worldbank.org/search/dataset/0038379

Climate effects on health incidence and behavior:

<https://www.oecd.org/en/data/datasets/oecd-health-statistics.html>

Solar radiation and mortality per week

[https://data-explorer.oecd.org/vis?fs[0]=Topic%2C1%7CHealth%23HEA%23%7CHealth%20status%23HEA\_STA%23&pg=0&fc=Topic&bp=true&snb=16&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD\_HEALTH\_MORTALITY%40DF\_MORTALITY&df[ag]=OECD.ELS.HD&df[vs]=1.0&dq=.W.M.\_T.\_T.&pd=2023-W01%2C&to[TIME\_PERIOD]=false](https://data-explorer.oecd.org/vis?fs%5B0%5D=Topic%2C1%7CHealth%23HEA%23%7CHealth%20status%23HEA_STA%23&pg=0&fc=Topic&bp=true&snb=16&df%5Bds%5D=dsDisseminateFinalDMZ&df%5Bid%5D=DSD_HEALTH_MORTALITY%40DF_MORTALITY&df%5Bag%5D=OECD.ELS.HD&df%5Bvs%5D=1.0&dq=.W.M._T._T.&pd=2023-W01%2C&to%5BTIME_PERIOD%5D=false)

Climate factors to predict mortality rate

Solar radiation per country and years of life lost due to Alzheimers disease

Other dataset:

Pharmaceutical use in dataset:

[https://data-explorer.oecd.org/vis?lc=en&df[ds]=dsDisseminateFinalDMZ&df[id]=HEALTH\_PHMC%40DF\_PHMC\_CONSUM&df[ag]=OECD.ELS.HD&df[vs]=latest&dq=....J01&pd=2010%2C&to[TIME\_PERIOD]=false](https://data-explorer.oecd.org/vis?lc=en&df%5Bds%5D=dsDisseminateFinalDMZ&df%5Bid%5D=HEALTH_PHMC%40DF_PHMC_CONSUM&df%5Bag%5D=OECD.ELS.HD&df%5Bvs%5D=latest&dq=....J01&pd=2010%2C&to%5BTIME_PERIOD%5D=false)

**2024-10-26: Next steps:**

POWER (This was resolved by new dataset, see below for details)

~~location by single point or aggregate of all points in a country (median)~~

* ~~Ask andre~~
* ~~Single point easier (chatgpt may be able to do it)~~
* ~~Retrieve data for all countries (Maybe single point better)~~
* ~~Calculate Year statistic from months (average and sum)~~

OECD

* Evaluate which data points we have per country (Years, and sex)
* Make a selection of years to use. Maybe 2013-2018? Do we do before pandemic?
* Countries with missing years? Do we remove, do we impute?
* Do we have male and female combined for all? if not calculate
* Do we have enough data male female separate to use in a plot? (optional)
* Do we aggregate year information, or select one year? Let’s start with the simplified case and choose the most complete year of info after 2020

**2024-10-29 Status update and Next Steps:**

* After adding country to new\_df, we discovered that the dataset only covers Portugal and Spain
  + We couldn’t find a convenient way to export data from all countries from the original source of the radiation data.
  + Because of this, we decided to use a [new 2020 Solar Potential dataset](https://datacatalog.worldbank.org/search/dataset/0038379) to represent radiation exposure by country for year 2020
  + Our original plan to compare solar radiation to years life lost remains the same
  + The version 1 python files were moved to [folder Analysis v1](https://drive.google.com/drive/folders/1VKcagkxdmC0bbDQxZpffFiILTPHMIjKx?usp=drive_link)
* Next steps:
  + Make new repo and share with team
  + Redo simple data exploration
  + Clean data
  + Join Solar Potential dataset with OECD years life lost using country code or country name