Introduction

National Pollutant Discharge Elimination System (NPDES)

- The NPDES permit program addresses water pollution by regulating point sources that discharge pollutants to waters of the United States.
- Home Site on EPA.GOV
- Data Completeness Dashboard
- Data Downloads Page
    - Toxics Release Inventory
        - Has many subjects - using Combined Releases, Releases-Complete, Facility Releases and Transfers, TRI Form Totals
        - 2020 and state of CO if possible - for Combined Releases """ elect distinct V_TRI_RELEASES_EZ.TRI_FACILITY_ID, V_TRI_RELEASES_EZ.FACILITY_NAME, V_TRI_RELEASES_EZ.DOC_CTRL_NUM, V_TRI_RELEASES_EZ.REPORTING_YEAR, V_TRI_RELEASES_EZ.CHEM_NAME, V_TRI_RELEASES_EZ.CAS_NUM, V_TRI_RELEASES_EZ.ELEMENTAL_METAL_INCLUDED, V_TRI_RELEASES_EZ.CATEGORY, V_TRI_RELEASES_EZ.REL_EST_AMT_SUM, V_TRI_RELEASES_EZ.CARCINOGEN, V_TRI_RELEASES_EZ.CLEAN_AIR, V_TRI_RELEASES_EZ.LIST_3350, V_TRI_RELEASES_EZ.STREET_ADDRESS, V_TRI_RELEASES_EZ.CITY_NAME, V_TRI_RELEASES_EZ.COUNTY_NAME, V_TRI_RELEASES_EZ.STATE_COUNTY_FIPS_CODE, V_TRI_RELEASES_EZ.STATE_ABBR, V_TRI_RELEASES_EZ.ZIP_CODE, V_TRI_RELEASES_EZ.REGION, V_TRI_RELEASES_EZ.FAC_CLOSED_IND, V_TRI_RELEASES_EZ.ASGN_AGENCY, V_TRI_RELEASES_EZ.ASGN_FEDERAL_IND, V_TRI_RELEASES_EZ.PARENT_CO_DB_NUM, V_TRI_RELEASES_EZ.PARENT_CO_NAME, V_TRI_RELEASES_EZ.STANDARDIZED_PARENT_COMPANY, V_TRI_RELEASES_EZ.EPA_REGISTRY_ID, V_TRI_RELEASES_EZ.TRADE_SECRET_IND, V_TRI_RELEASES_EZ.PRIMARY_SIC_CODE, V_TRI_RELEASES_EZ.INDUSTRY_CODE, V_TRI_RELEASES_EZ.SIC_CODES, V_TRI_RELEASES_EZ.NAICS_CODES, V_TRI_RELEASES_EZ.PRIMARY_NAICS_CODE, V_TRI_RELEASES_EZ.SRS_ID from V_TRI_RELEASES_EZ where (V_TRI_RELEASES_EZ.REPORTING_YEAR = '2020') and (V_TRI_RELEASES_EZ.STATE_ABBR = 'CO') """
        - CO - (couldn't do 2020, is this all) - for Releases-Complete """ Select distinct V_TRI_FORM_R_COMP_EZ.TRI_FACILITY_ID, V_TRI_FORM_R_COMP_EZ.FACILITY_NAME, V_TRI_FORM_R_COMP_EZ.EPA_REGISTRY_ID, V_TRI_FORM_R_COMP_EZ.STREET_ADDRESS, V_TRI_FORM_R_COMP_EZ.CITY_NAME, V_TRI_FORM_R_COMP_EZ.COUNTY_NAME, V_TRI_FORM_R_COMP_EZ.STATE_ABBR, V_TRI_FORM_R_COMP_EZ.ZIP_CODE, V_TRI_FORM_R_COMP_EZ.PRIMARY_SIC_CODE, V_TRI_FORM_R_COMP_EZ.SIC_CODES, V_TRI_FORM_R_COMP_EZ.PRIMARY_NAICS_CODE, V_TRI_FORM_R_COMP_EZ.NAICS_CODES, V_TRI_FORM_R_COMP_EZ.INDUSTRY_CODE,

```
V_TRI_FORM_R_COMP_EZ.DOC_CTRL_NUM, V_TRI_FORM_R_COMP_EZ.CHEM_NAME,
V_TRI_FORM_R_COMP_EZ.TRI_CHEM_ID,
V_TRI_FORM_R_COMP_EZ.ELEMENTAL_METAL_INCLUDED,
V_TRI_FORM_R_COMP_EZ.REPORTING_YEAR,
V_TRI_FORM_R_COMP_EZ.TOTAL_ON_SITE_RELEASE,
V_TRI_FORM_R_COMP_EZ.TOTAL_OFF_SITE_RELEASE,
V_TRI_FORM_R_COMP_EZ.TOTAL_ON_OFF_SITE_RELEASE,
V_TRI_FORM_R_COMP_EZ.AIR_TOTAL_RELEASE,
V_TRI_FORM_R_COMP_EZ.FUGITIVE_TOT_REL,
V_TRI_FORM_R_COMP_EZ.STACK_TOT_REL,
V_TRI_FORM_R_COMP_EZ.WATER_TOTAL_RELEASE,
V_TRI_FORM_R_COMP_EZ.WATER_BODY_NAME_1,
V_TRI_FORM_R_COMP_EZ.WATER_RELEASE_1,
V_TRI_FORM_R_COMP_EZ.STORM_WATER_PERCENT_1,
V_TRI_FORM_R_COMP_EZ.WATER_BODY_NAME_2,
V_TRI_FORM_R_COMP_EZ.WATER_RELEASE_2,
V_TRI_FORM_R_COMP_EZ.STORM_WATER_PERCENT_2,
V_TRI_FORM_R_COMP_EZ.WATER_BODY_NAME_3,
V_TRI_FORM_R_COMP_EZ.WATER_RELEASE_3,
V_TRI_FORM_R_COMP_EZ.STORM_WATER_PERCENT_3,
V_TRI_FORM_R_COMP_EZ.LAND_TOTAL_RELEASE,
V_TRI_FORM_R_COMP_EZ.RCRA_C_TOT_REL,
V_TRI_FORM_R_COMP_EZ.OTH_LANDF_TOT_REL,
V_TRI_FORM_R_COMP_EZ.SURF_IMP_TOT_REL,
V_TRI_FORM_R_COMP_EZ.LAND_TREA_TOT_REL,
V_TRI_FORM_R_COMP_EZ.OTH_DISP_TOT_REL,
V_TRI_FORM_R_COMP_EZ.OTHER_LAND_RELEASE,
V_TRI_FORM_R_COMP_EZ.UNINJ_TOTAL_RELEASE,
V_TRI_FORM_R_COMP_EZ.UNINJ_I_TOT_REL,
V_TRI_FORM_R_COMP_EZ.UNINJ_IIV_TOT_REL,
V_TRI_FORM_R_COMP_EZ.POTW_TOT_TRANSFER,
V_TRI_FORM_R_COMP_EZ.DISP_STORAGE_ONLY,
V_TRI_FORM_R_COMP_EZ.DISP_SOLID_STAB,
V_TRI_FORM_R_COMP_EZ.DISP_WASTEWATER_TREAT,
V_TRI_FORM_R_COMP_EZ.DISP_UNGRND_INJ,
V_TRI_FORM_R_COMP_EZ.DISP_LANDFILL_SURF_IMP,
V_TRI_FORM_R_COMP_EZ.DISP_LAND_TREATMENT,
V_TRI_FORM_R_COMP_EZ.DISP_OTHER_LAND_DISP,
V_TRI_FORM_R_COMP_EZ.DISP_TRANS_WASTE_BROKER,
V_TRI_FORM_R_COMP_EZ.SRS_ID,
V_TRI_FORM_R_COMP_EZ.WASTE_ROCK_MANAGED_PILE,
V_TRI_FORM_R_COMP_EZ.WASTE_ROCK_QUANTITY from V_TRI_FORM_R_COMP_EZ
where (V_TRI_FORM_R_COMP_EZ.STATE_ABBR = 'CO') """
```

- CO (2020 returned 0 rows, so took that filter out) for report 3 the Facility Releases and Transfers """ Select distinct V_TRI_REL_AND_TRAN_EZ.TRI_FACILITY_ID, V_TRI_REL_AND_TRAN_EZ.FACILITY_NAME, V_TRI_REL_AND_TRAN_EZ.DOC_CTRL_NUM, V_TRI_REL_AND_TRAN_EZ.REPORTING_YEAR, V_TRI_REL_AND_TRAN_EZ.CHEM_NAME, V_TRI_REL_AND_TRAN_EZ.CAS_NUM, V_TRI_REL_AND_TRAN_EZ.ELEMENTAL_METAL_INCLUDED, V_TRI_REL_AND_TRAN_EZ.CATEGORY, V_TRI_REL_AND_TRAN_EZ.REL_EST_AMT_SUM, V_TRI_REL_AND_TRAN_EZ.CARCINOGEN, V_TRI_REL_AND_TRAN_EZ.CLEAN_AIR, V_TRI_REL_AND_TRAN_EZ.LIST_3350, V_TRI_REL_AND_TRAN_EZ.STREET_ADDRESS, V_TRI_REL_AND_TRAN_EZ.CITY_NAME, V_TRI_REL_AND_TRAN_EZ.COUNTY_NAME, V_TRI_REL_AND_TRAN_EZ.STATE_COUNTY_FIPS_CODE, V_TRI_REL_AND_TRAN_EZ.STATE_ABBR, V_TRI_REL_AND_TRAN_EZ.ZIP_CODE, V_TRI_REL_AND_TRAN_EZ.REGION, V_TRI_REL_AND_TRAN_EZ.FAC_CLOSED_IND, V_TRI_REL_AND_TRAN_EZ.ASGN_AGENCY, V_TRI_REL_AND_TRAN_EZ.ASGN_FEDERAL_IND, V_TRI_REL_AND_TRAN_EZ.PARENT_CO_DB_NUM, V_TRI_REL_AND_TRAN_EZ.PARENT_CO_NAME, V_TRI_REL_AND_TRAN_EZ.STANDARDIZED_PARENT_COMPANY, V_TRI_REL_AND_TRAN_EZ.EPA_REGISTRY_ID, V_TRI_REL_AND_TRAN_EZ.TRADE_SECRET_IND, V_TRI_REL_AND_TRAN_EZ.PRIMARY_SIC_CODE, V_TRI_REL_AND_TRAN_EZ.INDUSTRY_CODE, V_TRI_REL_AND_TRAN_EZ.SIC_CODES, V_TRI_REL_AND_TRAN_EZ.NAICS_CODES, V_TRI_REL_AND_TRAN_EZ.PRIMARY_NAICS_CODE, V_TRI_REL_AND_TRAN_EZ.SRS_ID from V_TRI_REL_AND_TRAN_EZ where (V_TRI_REL_AND_TRAN_EZ.STATE_ABBR = 'CO') """

Read the Data

```
In [1]:  import pandas as pd
         #pd.set_option('display.max_columns', None)
         #pd.set_option('display.max_rows', None)
```

```
In [2]:  def desc_entity(ent_df) :
             print(f"Shape: {ent_df.shape}")
             print(f"\nRows/Index: {ent_df.index}")
             print(f"\nColumns: \n{ent_df.dtypes}")
             print("\n Sample Data: \n")
             print(ent_df.head())
```

```
In [ ]:  comb_rls_raw = pd.read_csv("Data/combined_releases.csv")
         desc_entity(comb_rls_raw)
```

```
In [ ]:
```

```
fac_rls_raw = pd.read_csv("Data/facility_releases_and_transfers.CSV")
desc_entity(fac_rls_raw)
```

In [3]:
```
rls_complt_raw = pd.read_csv("Data/releases_complete2.CSV")
desc_entity(rls_complt_raw)
```

Shape: (96976, 56)

Rows/Index: RangeIndex(start=0, stop=96976, step=1)

Columns:
TRI_FACILITY_ID   object
FACILITY_NAME   object
EPA_REGISTRY_ID   float64
STREET_ADDRESS   object
CITY_NAME   object
COUNTY_NAME   object
STATE_ABBR   object
ZIP_CODE   int64
PRIMARY_SIC_CODE   object
SIC_CODES   object
PRIMARY_NAICS_CODE   int64
NAICS_CODES   object
INDUSTRY_CODE   int64
DOC_CTRL_NUM   float64
CHEM_NAME   object
TRI_CHEM_ID   object
ELEMENTAL_METAL_INCLUDED   object
REPORTING_YEAR   int64
TOTAL_ON_SITE_RELEASE   float64
TOTAL_OFF_SITE_RELEASE   float64
TOTAL_ON_OFF_SITE_RELEASE   float64
AIR_TOTAL_RELEASE   float64
FUGITIVE_TOT_REL   float64
STACK_TOT_REL   float64
WATER_TOTAL_RELEASE   float64
WATER_BODY_NAME_1   object
WATER_RELEASE_1   float64
STORM_WATER_PERCENT_1   float64
WATER_BODY_NAME_2   object
WATER_RELEASE_2   float64
STORM_WATER_PERCENT_2   float64
WATER_BODY_NAME_3   object
WATER_RELEASE_3   float64
STORM_WATER_PERCENT_3   float64
LAND_TOTAL_RELEASE   float64
RCRA_C_TOT_REL   float64
OTH_LANDF_TOT_REL   float64
SURF_IMP_TOT_REL   float64
LAND_TREA_TOT_REL   float64
OTH_DISP_TOT_REL   float64
OTHER_LAND_RELEASE   float64
UNINJ_TOTAL_RELEASE   float64
UNINJ_I_TOT_REL   float64
UNINJ_IIV_TOT_REL   float64
POTW_TOT_TRANSFER   float64
DISP_STORAGE_ONLY   float64
DISP_SOLID_STAB   float64
DISP_WASTEWATER_TREAT   float64
DISP_UNGRND_INJ   float64
DISP_LANDFILL_SURF_IMP   float64
DISP_LAND_TREATMENT   float64
DISP_OTHER_LAND_DISP   float64
```

```
         DISP_TRANS_WASTE_BROKER        float64
         SRS_ID                         float64
         WASTE_ROCK_MANAGED_PILE        float64
         WASTE_ROCK_QUANTITY            float64
         dtype: object

          Sample Data:

            TRI_FACILITY_ID                              FACILITY_NAME  \
         0   85540PHLPS4521U                    FREEPORT-MCMORAN MORENCI INC
         1   84104CLSSC2587W                          CLASSIC CABINETS INC
         2   59806STNCNMULLA  SMURFIT-STONE CONTAINER ENTERPRISES INCMISSOUL...
         3   68008CRGLL650IN             CARGILL CORN MILLING NORTH AMERICA
         4   85043PHNXT5110W                      CHEVRON PHOENIX TERMINAL

            EPA_REGISTRY_ID              STREET_ADDRESS        CITY_NAME COUNTY_NAME  \
         0     1.100390e+11            4521 N US HWY 191         MORENCI    GREENLEE
         1              NaN       3045 WEST DIRECTORS ROW  SALT LAKE CITY   SALT LAKE
         2     1.100000e+11           14377 PULP MILL RD        MISSOULA    MISSOULA
         3     1.100600e+11   650 INDUSTRIAL PARK DRIVE          BLAIR  WASHINGTON
         4     1.100000e+11           5110 W MADISON ST         PHOENIX    MARICOPA

            STATE_ABBR    ZIP_CODE PRIMARY_SIC_CODE SIC_CODES  ...  DISP_SOLID_STAB  \
         0        AZ        85540              NaN       NaN  ...            0.0
         1        UT        84104             2434    2434NA  ...            0.0
         2        MT        59806              NaN       NaN  ...            0.0
         3        NE    680082649              NaN       NaN  ...            0.0
         4        AZ        85043              NaN       NaN  ...            0.0

            DISP_WASTEWATER_TREAT  DISP_UNGRND_INJ  DISP_LANDFILL_SURF_IMP  \
         0                   0.0              0.0                    0.00
         1                   0.0              0.0                    0.00
         2                   0.0              0.0                    2.36
         3                   0.0              0.0                    0.00
         4                   0.0              0.0                    0.00

            DISP_LAND_TREATMENT DISP_OTHER_LAND_DISP DISP_TRANS_WASTE_BROKER      SRS_ID  \
         0                 0.0                  0.0                     0.0  649699.0
         1                 0.0                  0.0                     0.0   25452.0
         2                 0.0                  0.0                     0.0   40576.0
         3                 0.0                  0.0                     0.0  149674.0
         4                 0.0                  0.0                     0.0    4754.0

            WASTE_ROCK_MANAGED_PILE  WASTE_ROCK_QUANTITY
         0                    NaN                  NaN
         1                    NaN                  NaN
         2                    NaN                  NaN
         3                    NaN                  NaN
         4                    NaN                  NaN

         [5 rows x 56 columns]
         C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
         eshell.py:3146: DtypeWarning: Columns (11) have mixed types.Specify dtype option on impo
         rt or set low_memory=False.
           has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

In [ ]:
```python
form_tot_raw = pd.read_csv("Data/TRI_Form_Totals.CSV")
desc_entity(form_tot_raw)
```

Analyze Combined Releases

```
In [ ]:   q1 = comb_rls_raw.groupby('FACILITY_NAME')['FACILITY_NAME'].count()
          filter = fn >= 5
          fac_nm_ge5 = q1[filter]
          print(fac_nm_ge5)
          #q3 = q2.unique()
          #q3
```

```
In [ ]:   #fac_nm_ge5.index
          filter = comb_rls_raw['FACILITY_NAME'].isin(fac_nm_ge5.index)
          comb_rls_sample = comb_rls_raw.loc[filter]
          print(comb_rls_sample.head())
          print(comb_rls_sample.shape)
```

```
In [ ]:   comb_rls_sample.dtypes
```

```
In [ ]:   columns = ['EXTERNAL_PERMIT_NMBR', 'FACILITY_NAME', 'STATE_CODE', ]
```

Analyze Releases Complete

Primary Key Analysis

- Primary Key Candidate, not quite - pk_candidate = rls_complt_gt0.groupby(['TRI_FACILITY_ID', 'DOC_CTRL_NUM', 'TRI_CHEM_ID', 'ELEMENTAL_METAL_INCLUDED'],
- Primary Key Candidate - Works - ['TRI_FACILITY_ID', 'DOC_CTRL_NUM', 'TRI_CHEM_ID', 'ELEMENTAL_METAL_INCLUDED', 'TOTAL_ON_OFF_SITE_RELEASE'],

Need to know which numbers subtotal to which numbers which then total to which?

```
In [4]:   print(rls_complt_raw.dtypes)
          print(rls_complt_raw.shape)
```

```
TRI_FACILITY_ID              object
FACILITY_NAME                object
EPA_REGISTRY_ID              float64
STREET_ADDRESS               object
CITY_NAME                    object
COUNTY_NAME                  object
STATE_ABBR                   object
ZIP_CODE                      int64
PRIMARY_SIC_CODE             object
SIC_CODES                    object
PRIMARY_NAICS_CODE            int64
NAICS_CODES                  object
INDUSTRY_CODE                 int64
DOC_CTRL_NUM                 float64
CHEM_NAME                    object
TRI_CHEM_ID                  object
ELEMENTAL_METAL_INCLUDED     object
REPORTING_YEAR                int64
TOTAL_ON_SITE_RELEASE        float64
TOTAL_OFF_SITE_RELEASE       float64
```

```
            TOTAL_ON_OFF_SITE_RELEASE       float64
            AIR_TOTAL_RELEASE               float64
            FUGITIVE_TOT_REL                float64
            STACK_TOT_REL                   float64
            WATER_TOTAL_RELEASE             float64
            WATER_BODY_NAME_1                object
            WATER_RELEASE_1                 float64
            STORM_WATER_PERCENT_1           float64
            WATER_BODY_NAME_2                object
            WATER_RELEASE_2                 float64
            STORM_WATER_PERCENT_2           float64
            WATER_BODY_NAME_3                object
            WATER_RELEASE_3                 float64
            STORM_WATER_PERCENT_3           float64
            LAND_TOTAL_RELEASE              float64
            RCRA_C_TOT_REL                  float64
            OTH_LANDF_TOT_REL               float64
            SURF_IMP_TOT_REL                float64
            LAND_TREA_TOT_REL               float64
            OTH_DISP_TOT_REL                float64
            OTHER_LAND_RELEASE              float64
            UNINJ_TOTAL_RELEASE             float64
            UNINJ_I_TOT_REL                 float64
            UNINJ_IIV_TOT_REL               float64
            POTW_TOT_TRANSFER               float64
            DISP_STORAGE_ONLY               float64
            DISP_SOLID_STAB                 float64
            DISP_WASTEWATER_TREAT           float64
            DISP_UNGRND_INJ                 float64
            DISP_LANDFILL_SURF_IMP          float64
            DISP_LAND_TREATMENT             float64
            DISP_OTHER_LAND_DISP            float64
            DISP_TRANS_WASTE_BROKER         float64
            SRS_ID                          float64
            WASTE_ROCK_MANAGED_PILE         float64
            WASTE_ROCK_QUANTITY             float64
            dtype: object
            (96976, 56)
```

In [5]:
```python
filter = rls_complt_raw['TOTAL_ON_OFF_SITE_RELEASE'] > 0
rls_complt_gt0 = rls_complt_raw[filter]
print(rls_complt_gt0.shape)
print(rls_complt_gt0.head(100))
```

```
            (96976, 56)
                TRI_FACILITY_ID                            FACILITY_NAME  \
            0     85540PHLPS4521U              FREEPORT-MCMORAN MORENCI INC
            1     84104CLSSC2587W                      CLASSIC CABINETS INC
            2     59806STNCNMULLA  SMURFIT-STONE CONTAINER ENTERPRISES INCMISSOUL...
            3     68008CRGLL650IN          CARGILL CORN MILLING NORTH AMERICA
            4     85043PHNXT5110W                 CHEVRON PHOENIX TERMINAL
            ..              ...                                       ...
            95    59044CNXRF803HI             CHS INC   LAUREL REFINERY
            96    85009CSCLT3010W               BAYSYSTEMS NORTH AMERICA
            97    81006PBLCS2005L  PUBLIC SERVICE CO OF COLORADO COMANCHE STATION
            98    85043MNRLF1832S                BORAL ROOFING - PHOENIX
            99    59101XXNBL700EX              EXXONMOBIL BILLINGS REFINERY


                EPA_REGISTRY_ID           STREET_ADDRESS      CITY_NAME   COUNTY_NAME  \
            0      1.100390e+11          4521 N US HWY 191        MORENCI      GREENLEE
            1               NaN    3045 WEST DIRECTORS ROW  SALT LAKE CITY    SALT LAKE
            2      1.100000e+11          14377 PULP MILL RD       MISSOULA      MISSOULA
            3      1.100600e+11  650 INDUSTRIAL PARK DRIVE          BLAIR    WASHINGTON
            4      1.100000e+11          5110 W MADISON ST        PHOENIX      MARICOPA
```

```
..         ...               ...               ...           ...
95     1.100010e+11         803 HWY 212 S        LAUREL   YELLOWSTONE
96     1.100000e+11       3010 W LINCOLN ST     PHOENIX     MARICOPA
97     1.100000e+11         2005 LIME RD        PUEBLO       PUEBLO
98     1.100000e+11       1832 S 51ST AVE       PHOENIX     MARICOPA
99     1.100000e+11      700 EXXONMOBIL RD      BILLINGS  YELLOWSTONE
```

```
    STATE_ABBR   ZIP_CODE PRIMARY_SIC_CODE SIC_CODES  ...  DISP_SOLID_STAB  \
0          AZ       85540             NaN       NaN  ...              0.0
1          UT       84104            2434    2434NA  ...              0.0
2          MT       59806             NaN       NaN  ...              0.0
3          NE   680082649             NaN       NaN  ...              0.0
4          AZ       85043             NaN       NaN  ...              0.0
..        ...         ...             ...       ... ...              ...
95         MT       59044            2911    2911NA  ...              0.0
96         AZ       85009            3086      3086  ...              0.0
97         CO       81006             NaN       NaN  ...              0.0
98         AZ       85043             NaN       NaN  ...              0.0
99         MT       59101            2911  29115171  ...              0.0
```

```
    DISP_WASTEWATER_TREAT  DISP_UNGRND_INJ  DISP_LANDFILL_SURF_IMP  \
0                     0.0              0.0                    0.00
1                     0.0              0.0                    0.00
2                     0.0              0.0                    2.36
3                     0.0              0.0                    0.00
4                     0.0              0.0                    0.00
..                    ...              ...                     ...
95                    0.0              0.0                    0.00
96                    0.0              0.0                    0.00
97                    0.0              0.0                    0.00
98                    0.0              0.0                   74.89
99                    0.0              0.0                    1.00
```

```
    DISP_LAND_TREATMENT DISP_OTHER_LAND_DISP DISP_TRANS_WASTE_BROKER      SRS_ID  \
0                   0.0                  0.0                     0.0  649699.0
1                   0.0                  0.0                     0.0   25452.0
2                   0.0                  0.0                     0.0   40576.0
3                   0.0                  0.0                     0.0  149674.0
4                   0.0                  0.0                     0.0    4754.0
..                  ...                  ...                     ...       ...
95                  0.0                  0.0                     0.0    5090.0
96                  0.0                  0.0                     0.0  650002.0
97                  0.0                  0.0                     0.0  152231.0
98                  0.0                  0.0                     0.0  149583.0
99                  0.0                  0.0                     0.0   26740.0
```

```
    WASTE_ROCK_MANAGED_PILE  WASTE_ROCK_QUANTITY
0                       NaN                  NaN
1                       NaN                  NaN
2                       NaN                  NaN
3                       NaN                  NaN
4                       NaN                  NaN
..                      ...                  ...
95                      NaN                  NaN
96                      NaN                  NaN
97                      NaN                  NaN
98                      NaN                  NaN
99                      NaN                  NaN
```

```
[100 rows x 56 columns]
```

In [6]:
```python
rls_complt_gt0['TOTAL_ON_OFF_SITE_RELEASE'].head(100)
```

```
0      530500.00
```

```
Out[6]: 1        10000.00
        2           3.20
        3           1.90
        4         184.00
                   ...
        95      17500.00
        96         10.00
        97      43000.00
        98         77.49
        99      28424.00
        Name: TOTAL_ON_OFF_SITE_RELEASE, Length: 100, dtype: float64
```

In [ ]:
```python
q1 = rls_complt_gt0.groupby(['FACILITY_NAME'])['FACILITY_NAME'].count()
print(q1)
#filter = fn >= 5
#fac_nm_ge5 = q1[filter]
#print(fac_nm_ge5)
```

In [28]:
```python
pk_candidate = rls_complt_gt0.groupby(['TRI_FACILITY_ID', 'DOC_CTRL_NUM', 'TRI_CHEM_ID'
                                       as_index=False).count()
#pk_candidate = rls_complt_raw.groupby(['EPA_REGISTRY_ID', 'DOC_CTRL_NUM', 'TRI_CHEM_ID
#pk_candidate =rls_complt_raw.groupby(['FACILITY_NAME', 'DOC_CTRL_NUM', 'TRI_CHEM_ID',
#                                      ['FACILITY_NAME', 'DOC_CTRL_NUM', 'TRI_CHEM_ID', '

#pk_candidate = rls_complt_raw.groupby(['TRI_FACILITY_ID', 'DOC_CTRL_NUM', 'TRI_CHEM_ID

#pk_candidate =rls_complt_raw.groupby\
#(['TRI_FACILITY_ID', 'DOC_CTRL_NUM', 'TRI_CHEM_ID', 'EPA_REGISTRY_ID', 'SRS_ID', 'REPO
# ['TRI_FACILITY_ID', 'DOC_CTRL_NUM', 'TRI_CHEM_ID', 'EPA_REGISTRY_ID', 'SRS_ID', 'REPO

print(pk_candidate.shape)
print(pk_candidate.head(5))
```

```
(7221, 56)
   TRI_FACILITY_ID   DOC_CTRL_NUM TRI_CHEM_ID ELEMENTAL_METAL_INCLUDED  \
0  59011STLLWPOBOX  1.318220e+12       67561                         N
1  59011STLLWPOBOX  1.318220e+12        N090                         N
2  59011STLLWPOBOX  1.318220e+12        N100                         N
3  59011STLLWPOBOX  1.318220e+12        N420                         N
4  59011STLLWPOBOX  1.318220e+12        N458                         N

   TOTAL_ON_OFF_SITE_RELEASE  FACILITY_NAME  EPA_REGISTRY_ID  STREET_ADDRESS  \
0                   95426.00              1                1               1
1                   55482.00              1                1               1
2                   29046.00              1                1               1
3                    3564.50              1                1               1
4                       3.22              1                1               1

   CITY_NAME  COUNTY_NAME  STATE_ABBR  ZIP_CODE  PRIMARY_SIC_CODE  SIC_CODES  \
0          1            1           1         1                 0          0
1          1            1           1         1                 0          0
2          1            1           1         1                 0          0
3          1            1           1         1                 0          0
4          1            1           1         1                 0          0

   PRIMARY_NAICS_CODE  NAICS_CODES  INDUSTRY_CODE  CHEM_NAME  REPORTING_YEAR  \
0                   1            1              1          1               1
1                   1            1              1          1               1
2                   1            1              1          1               1
3                   1            1              1          1               1
4                   1            1              1          1               1
```

```
    TOTAL_ON_SITE_RELEASE  TOTAL_OFF_SITE_RELEASE  AIR_TOTAL_RELEASE  \
0                       1                       1                  1
1                       1                       1                  1
2                       1                       1                  1
3                       1                       1                  1
4                       1                       1                  1

    FUGITIVE_TOT_REL  STACK_TOT_REL  WATER_TOTAL_RELEASE  WATER_BODY_NAME_1  \
0                  1              1                    1                  0
1                  1              1                    1                  0
2                  1              1                    1                  0
3                  1              1                    1                  0
4                  1              1                    1                  0

    WATER_RELEASE_1  STORM_WATER_PERCENT_1  WATER_BODY_NAME_2  WATER_RELEASE_2  \
0                1                      0                  0                0
1                1                      0                  0                0
2                1                      0                  0                0
3                1                      0                  0                0
4                1                      0                  0                0

    STORM_WATER_PERCENT_2  WATER_BODY_NAME_3  WATER_RELEASE_3  \
0                      0                  0                0
1                      0                  0                0
2                      0                  0                0
3                      0                  0                0
4                      0                  0                0

    STORM_WATER_PERCENT_3  LAND_TOTAL_RELEASE  RCRA_C_TOT_REL  \
0                      0                  1               1
1                      0                  1               1
2                      0                  1               1
3                      0                  1               1
4                      0                  1               1

    OTH_LANDF_TOT_REL  SURF_IMP_TOT_REL  LAND_TREA_TOT_REL  OTH_DISP_TOT_REL  \
0                  1                 1                  1                 1
1                  1                 1                  1                 1
2                  1                 1                  1                 1
3                  1                 1                  1                 1
4                  1                 1                  1                 1

    OTHER_LAND_RELEASE  UNINJ_TOTAL_RELEASE  UNINJ_I_TOT_REL  \
0                   1                    1                1
1                   1                    1                1
2                   1                    1                1
3                   1                    1                1
4                   1                    1                1

    UNINJ_IIV_TOT_REL  POTW_TOT_TRANSFER  DISP_STORAGE_ONLY  DISP_SOLID_STAB  \
0                  1                  1                  1                1
1                  1                  1                  1                1
2                  1                  1                  1                1
3                  1                  1                  1                1
4                  1                  1                  1                1

    DISP_WASTEWATER_TREAT  DISP_UNGRND_INJ  DISP_LANDFILL_SURF_IMP  \
0                      1                1                       1
1                      1                1                       1
2                      1                1                       1
3                      1                1                       1
4                      1                1                       1

    DISP_LAND_TREATMENT  DISP_OTHER_LAND_DISP  DISP_TRANS_WASTE_BROKER  SRS_ID  \
```

```
0              1               1                1    1
1              1               1                1    1
2              1               1                1    1
3              1               1                1    1
4              1               1                1    1
```

```
   WASTE_ROCK_MANAGED_PILE  WASTE_ROCK_QUANTITY
0                        1                    0
1                        1                    0
2                        1                    0
3                        1                    0
4                        1                    0
```

In [27]:
```python
f1 = rls_complt_gt0['TRI_FACILITY_ID'] == '82005SRFRC300VE'
f2 = rls_complt_gt0['DOC_CTRL_NUM'] == 1318220000000
q = rls_complt_gt0.loc[f1&f2]
q
```

Out[27]:

	TRI_FACILITY_ID	FACILITY_NAME	EPA_REGISTRY_ID	STREET_ADDRESS	CITY_NAME	COUNTY_N
19313	82005SRFRC300VE	US DOD USAF FE WARREN AFB (PART)	1.100420e+11	300 VESLE DR SUITE 600 90 CES CEIE	FE WARREN AFB	LAR
85967	82005SRFRC300VE	US DOD USAF FE WARREN AFB (PART)	1.100420e+11	300 VESLE DR SUITE 600 90 CES CEIE	FE WARREN AFB	LAR

In [29]:
```python
filter = pk_candidate['STATE_ABBR'] > 1
#print(filter[:100])
#print(any(filter))
nonpk = pk_candidate[filter]
print(nonpk.shape)
nonpk_nonna = nonpk.dropna()
print(nonpk_nonna)
print(nonpk_nonna.shape)
```

```
(0, 56)
Empty DataFrame
Columns: [TRI_FACILITY_ID, DOC_CTRL_NUM, TRI_CHEM_ID, ELEMENTAL_METAL_INCLUDED, TOTAL_ON
_OFF_SITE_RELEASE, FACILITY_NAME, EPA_REGISTRY_ID, STREET_ADDRESS, CITY_NAME, COUNTY_NAM
E, STATE_ABBR, ZIP_CODE, PRIMARY_SIC_CODE, SIC_CODES, PRIMARY_NAICS_CODE, NAICS_CODES, I
NDUSTRY_CODE, CHEM_NAME, REPORTING_YEAR, TOTAL_ON_SITE_RELEASE, TOTAL_OFF_SITE_RELEASE,
AIR_TOTAL_RELEASE, FUGITIVE_TOT_REL, STACK_TOT_REL, WATER_TOTAL_RELEASE, WATER_BODY_NAME
_1, WATER_RELEASE_1, STORM_WATER_PERCENT_1, WATER_BODY_NAME_2, WATER_RELEASE_2, STORM_WA
TER_PERCENT_2, WATER_BODY_NAME_3, WATER_RELEASE_3, STORM_WATER_PERCENT_3, LAND_TOTAL_REL
EASE, RCRA_C_TOT_REL, OTH_LANDF_TOT_REL, SURF_IMP_TOT_REL, LAND_TREA_TOT_REL, OTH_DISP_T
OT_REL, OTHER_LAND_RELEASE, UNINJ_TOTAL_RELEASE, UNINJ_I_TOT_REL, UNINJ_IIV_TOT_REL, POT
W_TOT_TRANSFER, DISP_STORAGE_ONLY, DISP_SOLID_STAB, DISP_WASTEWATER_TREAT, DISP_UNGRND_I
NJ, DISP_LANDFILL_SURF_IMP, DISP_LAND_TREATMENT, DISP_OTHER_LAND_DISP, DISP_TRANS_WASTE_
BROKER, SRS_ID, WASTE_ROCK_MANAGED_PILE, WASTE_ROCK_QUANTITY]
Index: []
(0, 56)
```

In [9]:
```python
# Not needed if 0 rows
nonpk_nonna.head()
```

Out[9]:

	TRI_FACILITY_ID	DOC_CTRL_NUM	TRI_CHEM_ID	ELEMENTAL_METAL_INCLUDED	FACILITY_NAME
444	59644GRYMN45MIL	1.318220e+12	7439921	N	2
445	59644GRYMN45MIL	1.318220e+12	N458	Y	2
446	59644GRYMN45MIL	1.319220e+12	N420	Y	2
447	59644GRYMN45MIL	1.319220e+12	N458	Y	2
593	68008CRGLL650IN	1.318220e+12	107028	N	2

5 rows × 56 columns

In [ ]:
```python
# Not needed if 0 rows
filter1 = rls_complt_raw['TRI_FACILITY_ID'] == '80913FRTCRCOMMA'
filter2 = rls_complt_raw['TRI_CHEM_ID'] == '7440508'
filter3 = rls_complt_raw['SRS_ID'] == 150086.0
filter4 = rls_complt_raw['EPA_REGISTRY_ID'] == 110000000000
filter5 = rls_complt_raw['REPORTING_YEAR'] == 2018
#filter3 = nonpk_nonna['SRS_ID'] == 149583.0
sample = rls_complt_raw.loc[filter1 & filter2 & filter3 & filter4 & filter5]
print(sample)
print(sample.shape)
```

In [ ]:

In [ ]:

In [ ]:
```python
rls_complt_raw
```

Transform the Data

Build a State-Year Aggregate from rls_complt_raw

In [6]:
```python
use_columns = ['REPORTING_YEAR', 'STATE_ABBR', 'TOTAL_ON_OFF_SITE_RELEASE', 'WATER_TOTA
st_yr = rls_complt_raw[use_columns].groupby(['REPORTING_YEAR', 'STATE_ABBR'], as_index=
st_yr
```

Out[6]:

	REPORTING_YEAR	STATE_ABBR	TOTAL_ON_OFF_SITE_RELEASE	WATER_TOTAL_RELEASE	LAND_TOTA
0	1987	AZ	6.813906e+07	3.010000e+03	4.9
1	1987	CO	3.571951e+07	3.329444e+06	1.3

	REPORTING_YEAR	STATE_ABBR	TOTAL_ON_OFF_SITE_RELEASE	WATER_TOTAL_RELEASE	LAND_TOTA
2	1987	MT	4.000236e+07	9.962220e+05	3.2
3	1987	NE	1.939952e+07	2.774398e+06	5.3
4	1987	NM	2.137732e+07	4.321000e+03	1.7
...	...	...	...	...	
226	2019	MT	6.177045e+07	9.854128e+04	5.9
227	2019	NE	1.784763e+07	5.514511e+06	3.5
228	2019	NM	1.662747e+07	1.416973e+05	1.4
229	2019	UT	1.990531e+08	1.433696e+05	1.8
230	2019	WY	1.850480e+07	8.441385e+03	1.5

231 rows × 5 columns

In [ ]:
```
TOTAL_ON_SITE_RELEASE, TOTAL_OFF_SITE_RELEASE, AIR_TOTAL_RELEASE, WATER_TOTAL_RELEASE,
LAND_TOTAL_RELEASE,
OTHER_LAND_RELEASE, OTH_DISP_TOT_REL, SURF_IMP_TOT_REL, OTH_LANDF_TOT_REL, TOTAL_ALL_OT
TOTAL_ALL_OTHER_CATEGS=LAND_TREA_TOT_REL, UNINJ_TOTAL_RELEASE, POTW_TOT_TRANSFER, DISP_


TOTAL_ON_SITE_RELEASE, TOTAL_OFF_SITE_RELEASE, AIR_TOTAL_RELEASE, FUGITIVE_TOT_REL, STA

WATER_TOTAL_RELEASE, WATER_BODY_NAME_1, WATER_RELEASE_1, STORM_WATER_PERCENT_1, WATER_B

LAND_TOTAL_RELEASE, RCRA_C_TOT_REL, OTH_LANDF_TOT_REL, SURF_IMP_TOT_REL, LAND_TREA_TOT_
```

In [6]:
```
filter=st_yr['REPORTING_YEAR'] == 2019
st_yr_2019 = st_yr.loc[filter]
st_yr_2019.shape
```

In [7]:
```
use_columns = ['REPORTING_YEAR', 'STATE_ABBR', 'TOTAL_ON_OFF_SITE_RELEASE', 'WATER_TOTA
st_yr = rls_complt_raw[use_columns].groupby(['REPORTING_YEAR', 'STATE_ABBR'], as_index=
st_yr
```

Out[7]:

	REPORTING_YEAR	STATE_ABBR	TOTAL_ON_OFF_SITE_RELEASE	WATER_TOTAL_RELEASE	LAND_TOTA
0	1987	AZ	6.813906e+07	3.010000e+03	4.9
1	1987	CO	3.571951e+07	3.329444e+06	1.3
2	1987	MT	4.000236e+07	9.962220e+05	3.2
3	1987	NE	1.939952e+07	2.774398e+06	5.3
4	1987	NM	2.137732e+07	4.321000e+03	1.7

	REPORTING_YEAR	STATE_ABBR	TOTAL_ON_OFF_SITE_RELEASE	WATER_TOTAL_RELEASE	LAND_TOTA
...	...	...	...	...	
226	2019	MT	6.177045e+07	9.854128e+04	5.9:
227	2019	NE	1.784763e+07	5.514511e+06	3.5
228	2019	NM	1.662747e+07	1.416973e+05	1.4
229	2019	UT	1.990531e+08	1.433696e+05	1.8·
230	2019	WY	1.850480e+07	8.441385e+03	1.5:

231 rows × 5 columns

In [19]:
```python
st_yr['STATE_ABBR'].unique()
```

Out[19]: array(['AZ', 'CO', 'MT', 'NE', 'NM', 'UT', 'WY'], dtype=object)

In [31]:

```
C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
eshell.py:3146: DtypeWarning: Columns (11) have mixed types.Specify dtype option on impo
rt or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Out[31]:

	REPORTING_YEAR	TOTAL_ON_OFF_SITE_RELEASE
0	1987	1990
1	1988	2221
2	1989	2398
3	1990	2446
4	1991	2407
5	1992	2330
6	1993	2279
7	1994	2172
8	1995	1994
9	1996	2023
10	1997	1928
11	1998	2816
12	1999	2767
13	2000	3000
14	2001	3287
15	2002	3260

	REPORTING_YEAR	TOTAL_ON_OFF_SITE_RELEASE
16	2003	3274
17	2004	3324
18	2005	3289
19	2006	3307
20	2007	3307
21	2008	3412
22	2009	3292
23	2010	3331
24	2011	3366
25	2012	3437
26	2013	3512
27	2014	3544
28	2015	3451
29	2016	3439
30	2017	3506
31	2018	3481
32	2019	3386

In [ ]:

In [ ]:

State Lat/Lon Reference Table

- Source

In [23]:
```python
st_lat_lon = pd.read_csv("Data/state_lat_lon.csv")
print(st_lat_lon.shape)
print(st_lat_lon.head())
```

```
(51, 4)
        State State_Abbr   Latitude   Longitude
0     Alabama         AL  32.806671  -86.791130
1      Alaska         AK  61.370716 -152.404419
2     Arizona         AZ  33.729759 -111.431221
3    Arkansas         AR  34.969704  -92.373123
4  California         CA  36.116203 -119.681564
```

In [ ]:

Join to add the lat/lon to the st_yr df

In [9]:
```python
st_yr_lat_lon = st_yr.merge(st_lat_lon, left_on='STATE_ABBR', right_on='State_Abbr')
st_yr_lat_lon
```

Out[9]:

	REPORTING_YEAR	STATE_ABBR	TOTAL_ON_OFF_SITE_RELEASE	WATER_TOTAL_RELEASE	LAND_TOTA
0	1987	AZ	6.813906e+07	3010.000	4.9:
1	1988	AZ	7.329892e+07	22120.000	5.4
2	1989	AZ	5.216736e+07	2260.000	3.8
3	1990	AZ	5.850271e+07	158.000	4.4
4	1991	AZ	7.477241e+07	32960.000	6.4:
...	...	...	...	...	
226	2015	WY	2.096629e+07	15336.630	1.7(
227	2016	WY	1.858622e+07	6287.139	1.5
228	2017	WY	2.020885e+07	8668.364	1.6(
229	2018	WY	2.142348e+07	5537.395	1.7
230	2019	WY	1.850480e+07	8441.385	1.5:

231 rows × 9 columns

In [10]:
```python
f2019 = st_yr_lat_lon['REPORTING_YEAR'] == 2019
st_yr_2019 = st_yr_lat_lon.loc[f2019]
st_yr_2019.to_csv("Data/st_yr_2019.csv")
```

Chemical Lists

In [4]:
```python
chem_pfa_df = pd.read_csv("Data/chem_pfa.csv")
print(chem_pfa_df.shape)
print(chem_pfa_df.head())
```

```
(172, 5)
        CASRN   \
0  68391-08-2
1  97659-47-7
2  68188-12-5
3   3825-26-1
4  68515-62-8

  TRI Chemical Name for Listed Per- and Polyfluoroalkyl Substances (PFAS)  \
0                    Alcohols, C8-14, γ-ω-perfluoro
1                    Alkenes, C8-14 α-, δ-ω-perfluoro
2                 Alkyl iodides, C4-20, γ-ω-perfluoro
3                      Ammonium perfluorooctanoate
4  1,4-Benzenedicarboxylic acid, dimethyl ester, ...
```

```
                                  Chemical Structure  De Minimis  \
       0                                        NaN         1.0
       1                                        NaN         1.0
       2                                        NaN         1.0
       3                                        NaN         1.0
       4  Reaction product mixture of variable structure         1.0

          MPOU Thresholds (lbs)
       0                    100
       1                    100
       2                    100
       3                    100
       4                    100
```

In [5]:
```python
chem_non_pfa_df = pd.read_csv("Data/chem_non_pfa.csv")
print(chem_non_pfa_df.shape)
print(chem_non_pfa_df.head())
```

```
(792, 7)
        CASRN Chemical Name  Chemical Structure De Minimis  \
0  71751-41-2      Abamectin                 NaN        1.0
1  30560-19-1       Acephate                 NaN        1.0
2     75-07-0   Acetaldehyde                 NaN        0.1
3     60-35-5       Acetamide                NaN        0.1
4     75-05-8   Acetonitrile                NaN        1.0

   MPOU Thresholds (lbs) Category Description Category Member
0         25,000/10,000                  NaN             NaN
1         25,000/10,000                  NaN             NaN
2         25,000/10,000                  NaN             NaN
3         25,000/10,000                  NaN             NaN
4         25,000/10,000                  NaN             NaN
```

Build the full size csv

In [10]:
```python
y2019 = pd.read_csv("Data/years/y2019.CSV")
y2019.shape
```

Out[10]: (78458, 56)

In [12]:
```python
y2018 = pd.read_csv("Data/years/y2018.CSV", error_bad_lines=False)
y2018.shape
```

```
b'Skipping line 3694: expected 56 fields, saw 67\nSkipping line 3696: expected 56 field
s, saw 67\nSkipping line 16369: expected 56 fields, saw 67\n'
b'Skipping line 32133: expected 56 fields, saw 67\n'
b'Skipping line 51423: expected 56 fields, saw 67\n'
b'Skipping line 67394: expected 56 fields, saw 67\nSkipping line 68708: expected 56 fiel
ds, saw 78\n'
```

Out[12]: (79953, 56)

In [ ]:
```python
y2017 = pd.read_csv("Data/years/y2017.CSV")
y2017.shape
```

In [3]:
```python
fname_prefix = 'Data/years/'
```

```python
ctr = 1
for year in range(1994, 2020) :
    fname = fname_prefix + 'y' + str(year) + '.CSV'
    tmp_df = pd.read_csv(fname, error_bad_lines=False)
    if ctr == 1 :
        full_df = tmp_df.copy()
    else :
        full_df = full_df.append(tmp_df)
    ctr += 1
full_df.shape
```

```
C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
eshell.py:3146: DtypeWarning: Columns (7) have mixed types.Specify dtype option on impor
t or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
eshell.py:3146: DtypeWarning: Columns (7,16) have mixed types.Specify dtype option on im
port or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
eshell.py:3146: DtypeWarning: Columns (7,11,16) have mixed types.Specify dtype option on
import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
b'Skipping line 19424: expected 56 fields, saw 67\nSkipping line 22658: expected 56 fiel
ds, saw 67\n'
C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
eshell.py:3146: DtypeWarning: Columns (7,8,16,31) have mixed types.Specify dtype option
on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
eshell.py:3146: DtypeWarning: Columns (11,16) have mixed types.Specify dtype option on i
mport or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
eshell.py:3146: DtypeWarning: Columns (16) have mixed types.Specify dtype option on impo
rt or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
eshell.py:3146: DtypeWarning: Columns (16,31) have mixed types.Specify dtype option on i
mport or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
b'Skipping line 14378: expected 56 fields, saw 67\n'
b'Skipping line 21598: expected 56 fields, saw 67\n'
b'Skipping line 51282: expected 56 fields, saw 67\nSkipping line 63497: expected 56 fiel
ds, saw 67\n'
b'Skipping line 74633: expected 56 fields, saw 67\n'
b'Skipping line 14648: expected 56 fields, saw 67\n'
b'Skipping line 52350: expected 56 fields, saw 67\nSkipping line 56622: expected 56 fiel
ds, saw 67\nSkipping line 64206: expected 56 fields, saw 67\nSkipping line 64369: expect
ed 56 fields, saw 67\n'
b'Skipping line 73505: expected 56 fields, saw 67\n'
b'Skipping line 2253: expected 56 fields, saw 67\nSkipping line 4370: expected 56 field
s, saw 67\nSkipping line 15541: expected 56 fields, saw 67\n'
b'Skipping line 36454: expected 56 fields, saw 67\n'
b'Skipping line 60566: expected 56 fields, saw 67\n'
b'Skipping line 77356: expected 56 fields, saw 67\nSkipping line 79176: expected 56 fiel
ds, saw 67\n'
b'Skipping line 3694: expected 56 fields, saw 67\nSkipping line 3696: expected 56 field
s, saw 67\nSkipping line 16369: expected 56 fields, saw 67\n'
b'Skipping line 32133: expected 56 fields, saw 67\n'
b'Skipping line 51423: expected 56 fields, saw 67\n'
b'Skipping line 67394: expected 56 fields, saw 67\nSkipping line 68708: expected 56 fiel
ds, saw 78\n'
```

Out[3]: (2216280, 56)

In [14]:

```python
new_df = full_df.rename(columns = {
        '_TRI_FORM_R_COMP_EZ.TRI_FACILITY_ID' : 'TRI_FACILITY_ID',
        'V_TRI_FORM_R_COMP_EZ.FACILITY_NAME' : 'FACILITY_NAME',
        'V_TRI_FORM_R_COMP_EZ.EPA_REGISTRY_ID' : 'EPA_REGISTRY_ID',
        'V_TRI_FORM_R_COMP_EZ.STREET_ADDRESS' : ' STREET_ADDRESS',
        'V_TRI_FORM_R_COMP_EZ.CITY_NAME' : 'CITY_NAME',
        'V_TRI_FORM_R_COMP_EZ.COUNTY_NAME' : ' COUNTY_NAME',
        'V_TRI_FORM_R_COMP_EZ.STATE_ABBR' : 'STATE_ABBR',
        'V_TRI_FORM_R_COMP_EZ.ZIP_CODE' : 'ZIP_CODE',
        'V_TRI_FORM_R_COMP_EZ.PRIMARY_SIC_CODE' : 'PRIMARY_SIC_CODE',
        'V_TRI_FORM_R_COMP_EZ.SIC_CODES' : 'SIC_CODES',
        'V_TRI_FORM_R_COMP_EZ.PRIMARY_NAICS_CODE' : 'PRIMARY_NAICS_CODE',
        'V_TRI_FORM_R_COMP_EZ.NAICS_CODES' : 'NAICS_CODES',
        'V_TRI_FORM_R_COMP_EZ.INDUSTRY_CODE' : 'INDUSTRY_CODE',
        'V_TRI_FORM_R_COMP_EZ.DOC_CTRL_NUM' : ' DOC_CTRL_NUM',
        'V_TRI_FORM_R_COMP_EZ.CHEM_NAME' : 'CHEM_NAME',
        'V_TRI_FORM_R_COMP_EZ.TRI_CHEM_ID' : 'TRI_CHEM_ID',
        'V_TRI_FORM_R_COMP_EZ.ELEMENTAL_METAL_INCLUDED' : 'ELEMENTAL_METAL_INCLUDED',
        'V_TRI_FORM_R_COMP_EZ.REPORTING_YEAR' : 'REPORTING_YEAR',
        'V_TRI_FORM_R_COMP_EZ.TOTAL_ON_SITE_RELEASE' : 'TOTAL_ON_SITE_RELEASE',
        'V_TRI_FORM_R_COMP_EZ.TOTAL_OFF_SITE_RELEASE' : 'TOTAL_OFF_SITE_RELEASE',
        'V_TRI_FORM_R_COMP_EZ.TOTAL_ON_OFF_SITE_RELEASE' : 'TOTAL_ON_OFF_SITE_RELEASE',
        'V_TRI_FORM_R_COMP_EZ.AIR_TOTAL_RELEASE' : 'AIR_TOTAL_RELEASE',
        'V_TRI_FORM_R_COMP_EZ.FUGITIVE_TOT_REL' : 'FUGITIVE_TOT_REL',
        'V_TRI_FORM_R_COMP_EZ.STACK_TOT_REL' : 'STACK_TOT_REL',
        'V_TRI_FORM_R_COMP_EZ.WATER_TOTAL_RELEASE' : 'WATER_TOTAL_RELEASE',
        'V_TRI_FORM_R_COMP_EZ.WATER_BODY_NAME_1' : 'WATER_BODY_NAME_1',
        'V_TRI_FORM_R_COMP_EZ.WATER_RELEASE_1' : 'WATER_RELEASE_1',
        'V_TRI_FORM_R_COMP_EZ.STORM_WATER_PERCENT_1' : 'STORM_WATER_PERCENT_1',
        'V_TRI_FORM_R_COMP_EZ.WATER_BODY_NAME_2' : 'WATER_BODY_NAME_2',
        'V_TRI_FORM_R_COMP_EZ.WATER_RELEASE_2' : 'WATER_RELEASE_2',
        'V_TRI_FORM_R_COMP_EZ.STORM_WATER_PERCENT_2' : 'STORM_WATER_PERCENT_2',
        'V_TRI_FORM_R_COMP_EZ.WATER_BODY_NAME_3' : 'WATER_BODY_NAME_3',
        'V_TRI_FORM_R_COMP_EZ.WATER_RELEASE_3' : 'WATER_RELEASE_3',
        'V_TRI_FORM_R_COMP_EZ.STORM_WATER_PERCENT_3' : 'STORM_WATER_PERCENT_3',
        'V_TRI_FORM_R_COMP_EZ.LAND_TOTAL_RELEASE' : 'LAND_TOTAL_RELEASE',
        'V_TRI_FORM_R_COMP_EZ.RCRA_C_TOT_REL' : 'RCRA_C_TOT_REL',
        'V_TRI_FORM_R_COMP_EZ.OTH_LANDF_TOT_REL' : 'OTH_LANDF_TOT_REL',
        'V_TRI_FORM_R_COMP_EZ.SURF_IMP_TOT_REL' : 'SURF_IMP_TOT_REL',
        'V_TRI_FORM_R_COMP_EZ.LAND_TREA_TOT_REL' : 'LAND_TREA_TOT_REL',
        'V_TRI_FORM_R_COMP_EZ.OTH_DISP_TOT_REL' : 'OTH_DISP_TOT_REL',
        'V_TRI_FORM_R_COMP_EZ.OTHER_LAND_RELEASE' : 'OTHER_LAND_RELEASE',
        'V_TRI_FORM_R_COMP_EZ.UNINJ_TOTAL_RELEASE' : 'UNINJ_TOTAL_RELEASE',
        'V_TRI_FORM_R_COMP_EZ.UNINJ_I_TOT_REL' : 'UNINJ_I_TOT_REL',
        'V_TRI_FORM_R_COMP_EZ.UNINJ_IIV_TOT_REL' : 'UNINJ_IIV_TOT_REL',
        'V_TRI_FORM_R_COMP_EZ.POTW_TOT_TRANSFER' : 'POTW_TOT_TRANSFER',
        'V_TRI_FORM_R_COMP_EZ.DISP_STORAGE_ONLY' : 'DISP_STORAGE_ONLY',
        'V_TRI_FORM_R_COMP_EZ.DISP_SOLID_STAB' : 'DISP_SOLID_STAB',
        'V_TRI_FORM_R_COMP_EZ.DISP_WASTEWATER_TREAT' : 'DISP_WASTEWATER_TREAT',
        'V_TRI_FORM_R_COMP_EZ.DISP_UNGRND_INJ' : 'DISP_UNGRND_INJ',
        'V_TRI_FORM_R_COMP_EZ.DISP_LANDFILL_SURF_IMP' : 'DISP_LANDFILL_SURF_IMP',
        'V_TRI_FORM_R_COMP_EZ.DISP_LAND_TREATMENT' : 'DISP_LAND_TREATMENT',
        'V_TRI_FORM_R_COMP_EZ.DISP_OTHER_LAND_DISP' : 'DISP_OTHER_LAND_DISP',
        'V_TRI_FORM_R_COMP_EZ.DISP_TRANS_WASTE_BROKER' : 'DISP_TRANS_WASTE_BROKER',
        'V_TRI_FORM_R_COMP_EZ.SRS_ID' : 'SRS_ID',
        'V_TRI_FORM_R_COMP_EZ.WASTE_ROCK_MANAGED_PILE' : 'WASTE_ROCK_MANAGED_PILE',
        'V_TRI_FORM_R_COMP_EZ.WASTE_ROCK_QUANTITY' : 'WASTE_ROCK_QUANTITY'
})
```

```
#full_df['TOT_ALL_OTH_CATEGS'] = full_df['UNINJ_TOTAL_RELEASE'] + full_df['POTW_TOT_TRA
#full_df.shape
```

In [11]:
```
new_df.columns
```

Out[11]:
```
Index(['TRI_FACILITY_ID', 'FACILITY_NAME', 'EPA_REGISTRY_ID',
       ' STREET_ADDRESS', 'CITY_NAME', ' COUNTY_NAME', 'STATE_ABBR',
       'ZIP_CODE', 'PRIMARY_SIC_CODE', 'SIC_CODES', 'PRIMARY_NAICS_CODE',
       'NAICS_CODES', 'INDUSTRY_CODE', ' DOC_CTRL_NUM', 'CHEM_NAME',
       'TRI_CHEM_ID', 'ELEMENTAL_METAL_INCLUDED', 'REPORTING_YEAR',
       'TOTAL_ON_SITE_RELEASE', 'TOTAL_OFF_SITE_RELEASE',
       'TOTAL_ON_OFF_SITE_RELEASE', 'AIR_TOTAL_RELEASE', 'FUGITIVE_TOT_REL',
       'STACK_TOT_REL', 'WATER_TOTAL_RELEASE', 'WATER_BODY_NAME_1',
       'WATER_RELEASE_1', 'STORM_WATER_PERCENT_1', 'WATER_BODY_NAME_2',
       'WATER_RELEASE_2', 'STORM_WATER_PERCENT_2', 'WATER_BODY_NAME_3',
       'WATER_RELEASE_3', 'STORM_WATER_PERCENT_3', 'LAND_TOTAL_RELEASE',
       'RCRA_C_TOT_REL', 'OTH_LANDF_TOT_REL', 'SURF_IMP_TOT_REL',
       'LAND_TREA_TOT_REL', 'OTH_DISP_TOT_REL', 'OTHER_LAND_RELEASE',
       'UNINJ_TOTAL_RELEASE', 'UNINJ_I_TOT_REL', 'UNINJ_IIV_TOT_REL',
       'POTW_TOT_TRANSFER', 'DISP_STORAGE_ONLY', 'DISP_SOLID_STAB',
       'DISP_WASTEWATER_TREAT', 'DISP_UNGRND_INJ', 'DISP_LANDFILL_SURF_IMP',
       'DISP_LAND_TREATMENT', 'DISP_OTHER_LAND_DISP',
       'DISP_TRANS_WASTE_BROKER', 'SRS_ID', 'WASTE_ROCK_MANAGED_PILE',
       'WASTE_ROCK_QUANTITY'],
      dtype='object')
```

In [16]:
```
new_df['TOT_ALL_OTH_CATEGS'] = new_df['LAND_TREA_TOT_REL'] + new_df['UNINJ_TOTAL_RELEAS
```

In [18]:
```
new_df.columns
new_df.shape
```

Out[18]:
```
(2216280, 57)
```

In [19]:
```
new_df.to_csv("Data/all_years_rows_cols.CSV")
```

Build full state-year data

In [3]:
```
all_years_rows_cols_df = pd.read_csv("Data/all_years_rows_cols.CSV")
```

```
C:\Users\MickC\Documents\virtual_envs\data_eng\lib\site-packages\IPython\core\interactiv
eshell.py:3146: DtypeWarning: Columns (8,9,10,12,17) have mixed types.Specify dtype opti
on on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

In [4]:
```
print(all_years_rows_cols_df.head())
print(all_years_rows_cols_df.shape)
print(all_years_rows_cols_df.columns)
```

```
   Unnamed: 0  TRI_FACILITY_ID          FACILITY_NAME  EPA_REGISTRY_ID  \
0           0  79905LPSRF6500T  WESTERN EL PASO REFINERY     1.100421e+11
1           1  52001SWSSV3510C             PRAIRIE FARMS     1.100004e+11
2           2  30117FHGHT421GA           QUAKER HOUGHTON     1.100004e+11
3           3  49507GRDSM1350S        LILLY INDUSTRIES INC              NaN
```

```
        4          4  35146GRDNR199MA        GAC SPRINGVILLE INC       1.100641e+11

              STREET_ADDRESS       CITY_NAME   COUNTY_NAME STATE_ABBR ZIP_CODE   \
    0  6500 TROWBRIDGE DR         EL PASO       EL PASO          TX    79905
    1    3510 CENTRAL AVE         DUBUQUE       DUBUQUE          IA    52001
    2      421 GARRETT ST      CARROLLTON       CARROLL          GA    30117
    3  1350 STEELE AVE SW    GRAND RAPIDS          KENT          MI    49507
    4  5496 US HIGHWAY 11     SPRINGVILLE      ST CLAIR          AL    35146

       PRIMARY_SIC_CODE  ... DISP_WASTEWATER_TREAT  DISP_UNGRND_INJ   \
    0              2911  ...                   0.0              0.0
    1              2026  ...                   0.0              0.0
    2              2992  ...                   0.0              0.0
    3              2851  ...                   0.0              0.0
    4              2952  ...                   0.0              0.0

       DISP_LANDFILL_SURF_IMP  DISP_LAND_TREATMENT  DISP_OTHER_LAND_DISP   \
    0                     0.0                  0.0                   0.0
    1                     0.0                  0.0                   0.0
    2                     0.0                  0.0                   0.0
    3                     0.0                  0.0                   0.0
    4                   250.0                  0.0                   0.0

       DISP_TRANS_WASTE_BROKER      SRS_ID WASTE_ROCK_MANAGED_PILE   \
    0                      0.0   46508.0                      NaN
    1                      0.0  152363.0                      NaN
    2                      0.0   27516.0                      NaN
    3                      0.0   24851.0                      NaN
    4                      0.0   85282.0                      NaN

       WASTE_ROCK_QUANTITY   TOT_ALL_OTH_CATEGS
    0                  NaN                  0.0
    1                  NaN                  0.0
    2                  NaN                250.0
    3                  NaN                  0.0
    4                  NaN                250.0

    [5 rows x 58 columns]
    (2216280, 58)
    Index(['Unnamed: 0', 'TRI_FACILITY_ID', 'FACILITY_NAME', 'EPA_REGISTRY_ID',
           ' STREET_ADDRESS', 'CITY_NAME', ' COUNTY_NAME', 'STATE_ABBR',
           'ZIP_CODE', 'PRIMARY_SIC_CODE', 'SIC_CODES', 'PRIMARY_NAICS_CODE',
           'NAICS_CODES', 'INDUSTRY_CODE', ' DOC_CTRL_NUM', 'CHEM_NAME',
           'TRI_CHEM_ID', 'ELEMENTAL_METAL_INCLUDED', 'REPORTING_YEAR',
           'TOTAL_ON_SITE_RELEASE', 'TOTAL_OFF_SITE_RELEASE',
           'TOTAL_ON_OFF_SITE_RELEASE', 'AIR_TOTAL_RELEASE', 'FUGITIVE_TOT_REL',
           'STACK_TOT_REL', 'WATER_TOTAL_RELEASE', 'WATER_BODY_NAME_1',
           'WATER_RELEASE_1', 'STORM_WATER_PERCENT_1', 'WATER_BODY_NAME_2',
           'WATER_RELEASE_2', 'STORM_WATER_PERCENT_2', 'WATER_BODY_NAME_3',
           'WATER_RELEASE_3', 'STORM_WATER_PERCENT_3', 'LAND_TOTAL_RELEASE',
           'RCRA_C_TOT_REL', 'OTH_LANDF_TOT_REL', 'SURF_IMP_TOT_REL',
           'LAND_TREA_TOT_REL', 'OTH_DISP_TOT_REL', 'OTHER_LAND_RELEASE',
           'UNINJ_TOTAL_RELEASE', 'UNINJ_I_TOT_REL', 'UNINJ_IIV_TOT_REL',
           'POTW_TOT_TRANSFER', 'DISP_STORAGE_ONLY', 'DISP_SOLID_STAB',
           'DISP_WASTEWATER_TREAT', 'DISP_UNGRND_INJ', 'DISP_LANDFILL_SURF_IMP',
           'DISP_LAND_TREATMENT', 'DISP_OTHER_LAND_DISP',
           'DISP_TRANS_WASTE_BROKER', 'SRS_ID', 'WASTE_ROCK_MANAGED_PILE',
           'WASTE_ROCK_QUANTITY', 'TOT_ALL_OTH_CATEGS'],
          dtype='object')
```

In [5]:
```python
use_columns = ['STATE_ABBR', 'REPORTING_YEAR', 'TOTAL_ON_SITE_RELEASE', 'TOTAL_OFF_SITE
               'OTH_LANDF_TOT_REL', 'SURF_IMP_TOT_REL', 'OTH_DISP_TOT_REL', 'OTHER_LAND
st_yr_df = all_years_rows_cols_df[use_columns].groupby(['REPORTING_YEAR', 'STATE_ABBR']
```

In [6]:
```python
print(st_yr_df.shape)
print(st_yr_df.head())
```

```
(1451, 13)
   REPORTING_YEAR STATE_ABBR  TOTAL_ON_SITE_RELEASE  TOTAL_OFF_SITE_RELEASE  \
0            1994         AK              6674391.0                     0.0
1            1994         AL             95668912.0              13208086.0
2            1994         AR             42611792.0               2619025.0
3            1994         AS                12255.0                     0.0
4            1994         AZ             38389763.0                299871.0

   TOTAL_ON_OFF_SITE_RELEASE  AIR_TOTAL_RELEASE  WATER_TOTAL_RELEASE  \
0                  6674391.0          5476055.0            1150330.0
1                108876998.0         89019740.0            2256128.0
2                 45230817.0         30237452.0            1019200.0
3                    12255.0            12255.0                  0.0
4                 38689634.0          6507317.0                 39.0

   LAND_TOTAL_RELEASE  OTH_LANDF_TOT_REL  SURF_IMP_TOT_REL  OTH_DISP_TOT_REL  \
0             48006.0                0.0               0.0           47566.0
1           4393044.0                0.0          113496.0            2721.0
2          11355140.0              360.0         1335825.0            4476.0
3                 0.0                0.0               0.0               0.0
4          31882407.0                0.0          323004.0        23800990.0

   OTHER_LAND_RELEASE  TOT_ALL_OTH_CATEGS
0             47816.0             39305.0
1           3707600.0           6758167.0
2            563980.0          12311978.0
3                 0.0                 0.0
4          31444275.0            721538.0
```

In [58]:
```python
years_df = st_yr_df.groupby('REPORTING_YEAR').sum()
years_df.to_csv('Data/years.CSV')
```

- Quantitative Fields TOTAL_ON_SITE_RELEASE, TOTAL_OFF_SITE_RELEASE, AIR_TOTAL_RELEASE, WATER_TOTAL_RELEASE, LAND_TOTAL_RELEASE, OTHER_LAND_RELEASE, OTH_DISP_TOT_REL, SURF_IMP_TOT_REL, OTH_LANDF_TOT_REL, TOTAL_ALL_OTHER_CATEGS TOTAL_ALL_OTHER_CATEGS=LAND_TREA_TOT_REL, UNINJ_TOTAL_RELEASE, POTW_TOT_TRANSFER, DISP_STORAGE_ONLY, DISP_SOLID_STAB, DISP_WASTEWATER_TREAT, DISP_UNGRND_INJ, DISP_LANDFILL_SURF_IMP, DISP_LAND_TREATMENT, DISP_OTHER_LAND_DISP, DISP_TRANS_WASTE_BROKER

Join in State Lat/Lon data

In [25]:
```python
st_lat_lon = pd.read_csv("Data/state_lat_lon.csv")
print(st_lat_lon.shape)
print(st_lat_lon.head())
```

```
(51, 4)
        State State_Abbr   Latitude   Longitude
0     Alabama         AL  32.806671  -86.791130
1      Alaska         AK  61.370716 -152.404419
2     Arizona         AZ  33.729759 -111.431221
3    Arkansas         AR  34.969704  -92.373123
4  California         CA  36.116203 -119.681564
```

In [26]:
```python
st_yr_lat_lon = st_yr_df.merge(st_lat_lon, left_on='STATE_ABBR', right_on='State_Abbr')
st_yr_lat_lon
```

Out[26]:

	REPORTING_YEAR	STATE_ABBR	TOTAL_ON_SITE_RELEASE	TOTAL_OFF_SITE_RELEASE	TOTAL_ON_OF
0	1994	AK	6.674391e+06	0.000000e+00	
1	1995	AK	6.834300e+06	6.030000e+03	
2	1996	AK	6.908783e+06	0.000000e+00	
3	1997	AK	4.616687e+06	1.507000e+03	
4	1998	AK	3.165413e+08	1.588700e+04	
...	...	...	...	...	
1321	2015	WY	1.966233e+07	1.298926e+06	
1322	2016	WY	1.757797e+07	9.888577e+05	
1323	2017	WY	1.879151e+07	1.417276e+06	
1324	2018	WY	2.015080e+07	1.272679e+06	
1325	2019	WY	1.758119e+07	9.236099e+05	

1326 rows × 17 columns

Add in the tooltip

In [19]:
```python
# pd.set_option('display.max_rows', None)
# pd.set_option('display.max_columns', None)
# pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)
```

In [54]:
```python
tooltip_fld = pd.Series([], name='tooltip_fld', dtype='str')
ctr = 1
idx = ctr - 1

for st_yr_row in st_yr_lat_lon.iterrows() :
    t_str = 'Total On/Off Site: ' + str(round(st_yr_row[1]['TOTAL_ON_OFF_SITE_RELEASE']
            chr(10) + chr(847) + chr(10) + \
            "Total On: " + str(round(st_yr_row[1]['TOTAL_ON_SITE_RELEASE'])) + \
            chr(10) + chr(847) + chr(10) + \
            "Total Off: " + str(round(st_yr_row[1]['TOTAL_OFF_SITE_RELEASE'])) + \
            chr(10) + chr(847) + chr(10) + \
            "Total Land: " + str(round(st_yr_row[1]['LAND_TOTAL_RELEASE'])) + \
            chr(10) + chr(847) + chr(10) + \
            "Total Water: " + str(round(st_yr_row[1]['WATER_TOTAL_RELEASE'])) + \
            chr(10) + chr(847) + chr(10) + \
            "Total Air: " + str(round(st_yr_row[1]['AIR_TOTAL_RELEASE']))
    tooltip_fld.at[idx] = t_str
    ctr += 1
    idx += 1
```

```
tooltip_fld
```

Out[54]:
```
0           Total On/Off Site: 6674391\n\nTotal On: 6674391\n\nTotal Off: 0\n\nTot
al Land: 48006\n\nTotal Water: 1150330\n\nTotal Air: 5476055
1           Total On/Off Site: 6840330\n\nTotal On: 6834300\n\nTotal Off: 6030\n\nTot
al Land: 484099\n\nTotal Water: 950617\n\nTotal Air: 5399584
2           Total On/Off Site: 6908783\n\nTotal On: 6908783\n\nTotal Off: 0\n\nTo
tal Land: 6435\n\nTotal Water: 1024160\n\nTotal Air: 5878188
3           Total On/Off Site: 4618194\n\nTotal On: 4616687\n\nTotal Off: 1507\n\nT
otal Land: 2378\n\nTotal Water: 334183\n\nTotal Air: 4280126
4      Total On/Off Site: 316557156\n\nTotal On: 316541269\n\nTotal Off: 15887\n\nTotal
Land: 313420271\n\nTotal Water: 97777\n\nTotal Air: 3023221
                                                              ...
1321    Total On/Off Site: 20961261\n\nTotal On: 19662334\n\nTotal Off: 1298926\n\nTota
l Land: 17664777\n\nTotal Water: 15336\n\nTotal Air: 1982221
1322    Total On/Off Site: 18566826\n\nTotal On: 17577969\n\nTotal Off: 988858\n\nTot
al Land: 15466528\n\nTotal Water: 6287\n\nTotal Air: 2105154
1323    Total On/Off Site: 20208787\n\nTotal On: 18791510\n\nTotal Off: 1417276\n\nTot
al Land: 16610836\n\nTotal Water: 8668\n\nTotal Air: 2172006
1324    Total On/Off Site: 21423478\n\nTotal On: 20150798\n\nTotal Off: 1272679\n\nTot
al Land: 17729908\n\nTotal Water: 5537\n\nTotal Air: 2415353
1325    Total On/Off Site: 18504798\n\nTotal On: 17581188\n\nTotal Off: 923610\n\nTot
al Land: 15348504\n\nTotal Water: 8441\n\nTotal Air: 2224243
Name: tooltip_fld, Length: 1326, dtype: object
```

REPORTING_YEAR 1994 STATE_ABBR AK TOTAL_ON_SITE_RELEASE 6674391.0 TOTAL_OFF_SITE_RELEASE 0.0 TOTAL_ON_OFF_SITE_RELEASE 6674391.0 AIR_TOTAL_RELEASE 5476055.0 WATER_TOTAL_RELEASE 1150330.0 LAND_TOTAL_RELEASE 48006.0 OTH_LANDF_TOT_REL 0.0 SURF_IMP_TOT_REL 0.0 OTH_DISP_TOT_REL 47566.0 OTHER_LAND_RELEASE 47816.0 TOT_ALL_OTH_CATEGS 39305.0 State Alaska State_Abbr AK Latitude 61.370716 Longitude -152.404419

In [55]:
```python
st_yr_lat_lon_tt = st_yr_lat_lon.join(tooltip_fld)
```

In [56]:
```python
st_yr_lat_lon_tt.to_csv("Data/st_yr_lat_lon_tt.csv")
```

In [ ]:

In [ ]: