

# California Grants Dataset EDA

## EDA Using Python

- understand data
  - many columns are not very useful
  - some that may be useful are self reported, having no consistency in formatting between rows (award period, estimated amounts.)
- clean data
- analyze variables

```
In [2]: # import necessary packages
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [3]: grants_raw = pd.read_csv("ca_grants.csv")
```

```
In [4]: grants_raw.head()
```

```
Out[4]:
```

	PortalID	GrantID	Status	LastUpdated	ChangeNotes	AgencyDept	Title	Type	LOI
0	6481	DOJ-PROP56-2022-23-1	active	2022-07-18 17:28:34	Updated eligibility, suggested activities, and...	Department of Justice (Office of the Attorney ...	Tobacco Grant Program FY 2022-23 Request for P...	Grant	No
1	11966	NaN	active	2022-07-15 22:20:56	Application open date: July 15, 2022. Updated...	Department of Health Care Access and Information	California State Loan Repayment Program (SLRP)	Grant	No
2	11960	NaN	active	2022-07-14 22:35:54	NaN	CA Arts Council	Administering Organization – Individual Artist...	Grant	No
3	11957	NaN	active	2022-07-14 22:15:54	NaN	CA Arts Council	Administering Organization – Arts Administrato...	Grant	No
4	11912	NaN	active	2022-07-14 17:13:34	NaN	Department of Pesticide Regulation	Department of Pesticide Regulation 2023 Allian...	Grant	No

5 rows × 36 columns

```
In [5]: print(grants_raw.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 725 entries, 0 to 724
Data columns (total 36 columns):
#   Column                Non-Null Count  Dtype
---  -
0   PortalID              725 non-null    int64
1   GrantID               132 non-null    object
2   Status                725 non-null    object
3   LastUpdated          725 non-null    object
4   ChangeNotes          419 non-null    object
5   AgencyDept           725 non-null    object
6   Title                725 non-null    object
7   Type                 725 non-null    object
8   LOI                  724 non-null    object
9   Categories            725 non-null    object
10  CategorySuggestion   94 non-null     object
11  Purpose               724 non-null    object
12  Description           725 non-null    object
13  ApplicantType        721 non-null    object
14  ApplicantTypeNotes   604 non-null    object
15  Geography             557 non-null    object
16  FundingSource        721 non-null    object
17  FundingSourceNotes   523 non-null    object
18  MatchingFunds        725 non-null    object
19  MatchingFundsNotes   279 non-null    object
20  EstAvailFunds        713 non-null    object
21  EstAwards            725 non-null    object
22  EstAmounts           725 non-null    object
23  FundingMethod        719 non-null    object
24  FundingMethodNotes   426 non-null    object
25  OpenDate             720 non-null    object
26  ApplicationDeadline  715 non-null    object
27  AwardPeriod          712 non-null    object
28  ExpAwardDate         713 non-null    object
29  ElecSubmission       609 non-null    object
30  GrantURL             725 non-null    object
31  AgencyURL            724 non-null    object
32  AgencySubscribeURL   400 non-null    object
33  GrantEventsURL       275 non-null    object
34  ContactInfo          725 non-null    object
35  AwardStats           610 non-null    object
dtypes: int64(1), object(35)
memory usage: 204.0+ KB
None
```

```
In [6]: # Change ID of the grant to a categorical variable
grants_raw['PortalID'] = grants_raw['PortalID'].astype('object')
```

Removal of unnecessary columns

- columns with excessive missing values
- redundant information
- information that cannot realistically be useful or analyzed

```
In [7]: grants = grants_raw.drop(grants_raw.columns[[1,4,10,14,15,17,19,24,27,29,31,32,33,34,35])
grants.head()
```

Out[7]:

	PortalID	Status	LastUpdated	AgencyDept	Title	Type	LOI	Categories	Purpose
0	6481	active	2022-07-18 17:28:34	Department of Justice (Office of the Attorney ...	Tobacco Grant Program FY 2022-23 Request for P...	Grant	No	Education; Law, Justice, and Legal Services	The purpo: of this gra offere through the
1	11966	active	2022-07-15 22:20:56	Department of Health Care Access and Information	California State Loan Repayment Program (SLRP)	Grant	No	Health & Human Services	The Californ State Loā Repayme Program (S
2	11960	active	2022-07-14 22:35:54	CA Arts Council	Administering Organization – Individual Artist...	Grant	No	Disadvantaged Communities; Libraries and Arts	Th Administerir Organization Individual Ar
3	11957	active	2022-07-14 22:15:54	CA Arts Council	Administering Organization – Arts Administrato...	Grant	No	Disadvantaged Communities; Education; Employme...	The Ar Administrato Pipelir Fellowship pr
4	11912	active	2022-07-14 17:13:34	Department of Pesticide Regulation	Department of Pesticide Regulation 2023 Allian...	Grant	No	Agriculture; Disadvantaged Communities; Educat...	To promo safer, mo sustainab pest manage

5 rows × 21 columns



```
In [8]: # Columns we are left with
print(grants.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 725 entries, 0 to 724
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   PortalID              725 non-null    object
1   Status                725 non-null    object
2   LastUpdated          725 non-null    object
3   AgencyDept           725 non-null    object
4   Title                725 non-null    object
5   Type                 725 non-null    object
6   LOI                  724 non-null    object
7   Categories           725 non-null    object
8   Purpose              724 non-null    object
9   Description          725 non-null    object
10  ApplicantType        721 non-null    object
11  FundingSource        721 non-null    object
12  MatchingFunds       725 non-null    object
13  EstAvailFunds       713 non-null    object
14  EstAwards            725 non-null    object
15  EstAmounts          725 non-null    object
16  FundingMethod        719 non-null    object
17  OpenDate             720 non-null    object
18  ApplicationDeadline  715 non-null    object
19  ExpAwardDate        713 non-null    object
20  GrantURL             725 non-null    object
dtypes: object(21)
memory usage: 119.1+ KB
None
```

```
In [9]: grants.AgencyDept.value_counts()
grants.AgencyDept.unique()
```

```

Out[9]: array(['Department of Justice (Office of the Attorney General)',
'Department of Health Care Access and Information',
'CA Arts Council', 'Department of Pesticide Regulation',
"Governor's Office of Emergency Services",
'Department of Social Services',
'Department of Forestry and Fire Protection',
'Board of State and Community Corrections', 'CA State Library',
'Workforce Development Board',
'Department of Parks and Recreation',
"Governor's Office of Business and Economic Development",
'Coachella Valley Mountains Conservancy',
'Department of Conservation', 'Coastal Commission',
'Department of Resources Recycling and Recovery',
'CA Department of Food and Agriculture', 'CA Energy Commission',
'CA Natural Resources Agency', 'Victim Compensation Board',
'Department of Community Services and Development',
'Department of Housing and Community Development',
'Sierra Nevada Conservancy', 'Department of Water Resources',
'Employment Development Department', 'Strategic Growth Council',
'Wildlife Conservation Board', 'San Diego River Conservancy',
'Department of Transportation', 'Ocean Protection Council',
'Coastal Conservancy', 'Department of Fish and Wildlife',
'State Water Resources Control Board', 'Air Resources Board',
'Department of General Services', 'CA Volunteers',
'Transportation Commission', 'Department of Public Health',
'CA Business, Consumer Services, and Housing Agency',
'Department of Financial Protection and Innovation',
"State Treasurer's Office",
'Department of Alcoholic Beverage Control',
'Employment Training Panel', 'Department of Industrial Relations',
"Governor's Office of Planning and Research",
'Commission on the Status of Women and Girls',
'California Highway Patrol',
'Department of Toxic Substances Control',
'CA Department of Corrections and Rehabilitation',
'CA State Transportation Agency',
'CA Department of Veterans Affairs',
'Department of Cannabis Control', 'Department of Rehabilitation',
'Sacramento-San Joaquin Delta Conservancy',
'Student Aid Commission', 'Tahoe Conservancy',
'Public Utilities Commission',
'Infrastructure and Economic Development Bank',
'Commission on Peace Officer Standards and Training',
'CA Natural Resources Agency; Coastal Conservancy',
'CA Environmental Protection Agency; State Water Resources Control Board'],
dtype=object)

```

```

In [10]: grants.describe(include=["object"]) #no duplicate rows

```

Out[10]:

	PortalID	Status	LastUpdated	AgencyDept	Title	Type	LOI	Categories	Purpose	
<b>count</b>	725	725	725	725	725	725	724	725	724	
<b>unique</b>	725	3	725	61	659	3	2	196	628	
<b>top</b>	6481	closed	2022-07-18 17:28:34	Governor's Office of Emergency Services	Vertebrate Pest Control Research Program	Grant	No	Environment & Water	The purpose of the Cannabis Tax Fund Grant Pro...	P
<b>freq</b>	1	572	1	48	4	657	576	94	7	

4 rows × 21 columns

Converting the columns 'EstAwards', 'EstAmounts', and 'EstAvailFunds' into numeric variables. Unique values reveal that the entries for these two columns are formatted consistently. As many entries contain a range of values these columns were each split into 2, selecting their maximum and minimum values. Undeclared entries were replaced with a missing value (NaN).

In [11]: `grants['EstAwards'].unique()`

Out[11]:

```
array(['Dependant on number of submissions received, application process, etc.',
      'Exactly 1', 'Between 1 and 30', 'Between 1 and 63',
      'Between 1 and 6', 'Exactly 21', 'Exactly 10', 'Between 1 and 2',
      'Between 50 and 60', 'Between 4 and 7', 'Exactly 13', 'Exactly 2',
      'Exactly 5', 'Exactly 14', 'Exactly 20', 'Exactly 4', 'Exactly 3',
      'Between 35 and 35', 'Between 25 and 40', 'Between 2 and 10',
      'Between 500 and 1000', 'Between 30 and 35', 'Between 0 and 0',
      'Between 0 and 77', 'Exactly 7', 'Exactly 50', 'Between 10 and 20',
      'Between 0 and 100', 'Exactly 40', 'Between 0 and 20',
      'Between 1 and 8', 'Exactly 70', 'Exactly 6', 'Between 8 and 12',
      'Exactly 100', 'Between 0 and 62732', 'Between 0 and 270',
      'Between 0 and 675', 'Between 15 and 30', 'Exactly 225',
      'Between 2 and 4', 'Between 0 and 58', 'Between 1 and 17',
      'Between 20 and 30', 'Between 1 and 18',
      'Between 100000 and 500000', 'Between 1 and 10',
      'Between 60 and 79', 'Exactly 16', 'Between 1 and 12',
      'Between 65 and 85', 'Between 40 and 50', 'Between 16 and 30',
      'Between 1 and 5', 'Exactly 9', 'Between 1 and 4', 'Exactly 90',
      'Between 5 and 12', 'Exactly 8', 'Between 5 and 7',
      'Between 10 and 12', 'Between 3 and 5', 'Between 1 and 15',
      'Between 1 and 3', 'Between 0 and 3', 'Exactly 12', 'Exactly 432'],
      dtype=object)
```

In [12]:

```
awards = grants['EstAwards']
maxaward = []
minaward = []
for i in (range(len(awards))):
    if awards[i][0] == 'E':
        maxaward.append(int(''.join(filter(str.isdigit, awards[i]))))
        minaward.append(int(''.join(filter(str.isdigit, awards[i]))))
    elif awards[i][0] == 'B':
        maxaward.append(int(''.join(filter(str.isdigit, awards[i].rpartition('a')[2]))))
```

```

        minaward.append(int(''.join(filter(str.isdigit, awards[i].rpartition('a')[0])))
    else:
        maxaward.append(float('nan'))
        minaward.append(float('nan'))

amounts = grants['EstAmounts']
maxamnt = []
minamnt = []
for i in range(len(amounts)):
    if amounts[i][0] == 'E':
        maxamnt.append(int(''.join(filter(str.isdigit, amounts[i])))
        minamnt.append(int(''.join(filter(str.isdigit, amounts[i])))
    elif amounts[i][0] == 'B':
        maxamnt.append(int(''.join(filter(str.isdigit, amounts[i].rpartition('a')[2])))
        minamnt.append(int(''.join(filter(str.isdigit, amounts[i].rpartition('a')[0])))
    else:
        maxamnt.append(float('nan'))
        minamnt.append(float('nan'))

```

```

In [13]: grants['MaxAwards'] = maxaward
grants['MinAwards'] = minaward
grants = grants.drop('EstAwards', axis = 1)

grants['MaxAmounts'] = maxamnt
grants['MinAmounts'] = minamnt
grants = grants.drop('EstAmounts', axis = 1)

```

```

In [14]: availfunds = []
for i in range(len(grants['EstAvailFunds'])):
    if type(grants['EstAvailFunds'][i]) != str:
        availfunds.append(float('nan'))
    else:
        availfunds.append(int(''.join(filter(str.isdigit, grants['EstAvailFunds'][i])))

```

```

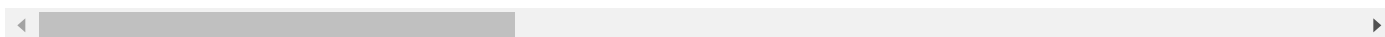
In [15]: grants['EstAvailFunds'] = availfunds
grants.head()

```

Out[15]:

	PortalID	Status	LastUpdated	AgencyDept	Title	Type	LOI	Categories	Purpose
0	6481	active	2022-07-18 17:28:34	Department of Justice (Office of the Attorney ...	Tobacco Grant Program FY 2022-23 Request for P...	Grant	No	Education; Law, Justice, and Legal Services	The purpose of this grant is to offer... through the
1	11966	active	2022-07-15 22:20:56	Department of Health Care Access and Information	California State Loan Repayment Program (SLRP)	Grant	No	Health & Human Services	The California State Loan Repayment Program (S
2	11960	active	2022-07-14 22:35:54	CA Arts Council	Administering Organization – Individual Artist...	Grant	No	Disadvantaged Communities; Libraries and Arts	The Administering Organization Individual Ar
3	11957	active	2022-07-14 22:15:54	CA Arts Council	Administering Organization – Arts Administrato...	Grant	No	Disadvantaged Communities; Education; Employme...	The Ar Administrato Pipelin Fellowship pr
4	11912	active	2022-07-14 17:13:34	Department of Pesticide Regulation	Department of Pesticide Regulation 2023 Allian...	Grant	No	Agriculture; Disadvantaged Communities; Educat...	To promote safer, more sustainable pest manage

5 rows x 23 columns



```
In [16]: print(grants.info()) # Our new columns are left with mostly missing values as a majori
```



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 725 entries, 0 to 724
Data columns (total 23 columns):
#   Column                Non-Null Count  Dtype
---  -
0   PortalID              725 non-null    object
1   Status                725 non-null    object
2   LastUpdated           725 non-null    object
3   AgencyDept            725 non-null    object
4   Title                 725 non-null    object
5   Type                  725 non-null    object
6   LOI                   724 non-null    object
7   Categories             725 non-null    object
8   Purpose               724 non-null    object
9   Description            725 non-null    object
10  ApplicantType         721 non-null    object
11  FundingSource         721 non-null    object
12  MatchingFunds        725 non-null    object
13  EstAvailFunds        713 non-null    float64
14  FundingMethod         719 non-null    object
15  OpenDate              720 non-null    object
16  ApplicationDeadline   715 non-null    object
17  ExpAwardDate         713 non-null    object
18  GrantURL              725 non-null    object
19  MaxAwards             148 non-null    float64
20  MinAwards             148 non-null    float64
21  MaxAmounts           231 non-null    float64
22  MinAmounts           231 non-null    float64
dtypes: float64(5), object(18)
memory usage: 130.4+ KB
None
```

In [17]: `grants.describe(include=["object"])`

Out[17]:

	PortalID	Status	LastUpdated	AgencyDept	Title	Type	LOI	Categories	Purpose	D
<b>count</b>	725	725	725	725	725	725	724	725	724	
<b>unique</b>	725	3	725	61	659	3	2	196	628	
<b>top</b>	6481	closed	2022-07-18 17:28:34	Governor's Office of Emergency Services	Vertebrate Pest Control Research Program	Grant	No	Environment & Water	The purpose of the Cannabis Tax Fund Grant Pro...	P
<b>freq</b>	1	572	1	48	4	657	576	94	7	

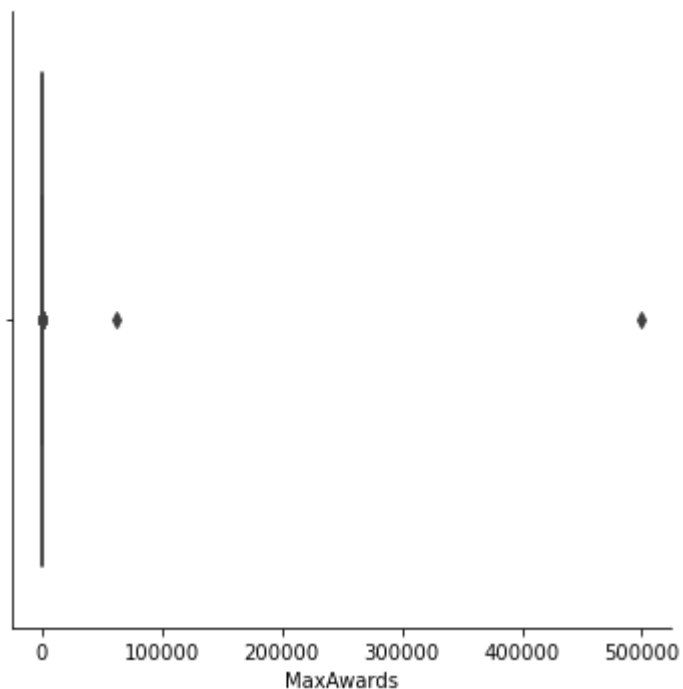
In [18]: `grants.describe()`

```
Out[18]:
```

	EstAvailFunds	MaxAwards	MinAwards	MaxAmounts	MinAmounts
<b>count</b>	7.130000e+02	148.000000	148.000000	2.310000e+02	2.310000e+02
<b>mean</b>	6.380992e+07	3834.182432	693.114865	4.843356e+07	3.098185e+05
<b>std</b>	3.500748e+08	41384.329800	8218.712299	4.652575e+08	1.404581e+06
<b>min</b>	1.000000e+00	0.000000	0.000000	1.380000e+02	0.000000e+00
<b>25%</b>	1.170000e+06	2.000000	1.000000	1.000000e+05	0.000000e+00
<b>50%</b>	5.000000e+06	7.000000	2.000000	3.500000e+05	5.000000e+03
<b>75%</b>	2.000000e+07	20.000000	12.250000	1.500000e+06	5.000000e+04
<b>max</b>	5.000000e+09	500000.000000	100000.000000	5.000000e+09	1.500000e+07

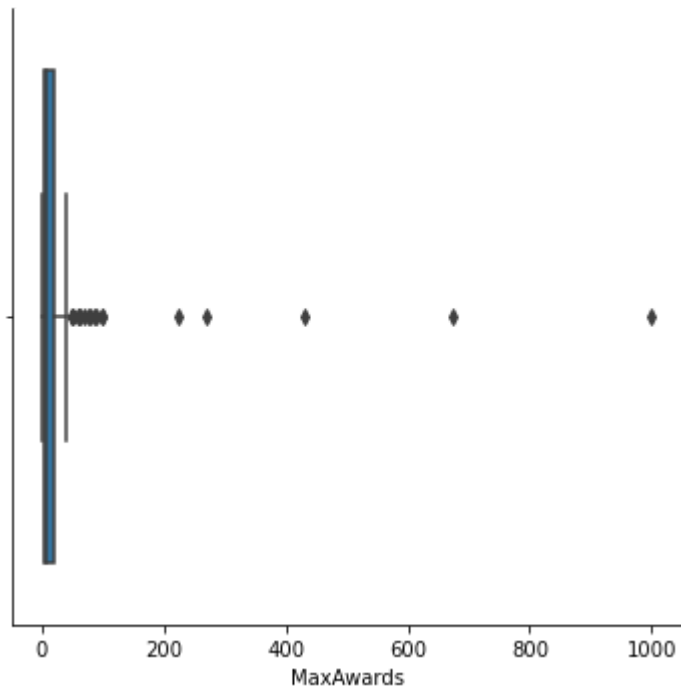
### Exploring the Numeric Variables

```
In [19]: sns.catplot(x = 'MaxAwards', kind = 'box', data = grants)
grants2 = grants[grants["MaxAwards"] < 50000] # remove the excessively large outliers
```



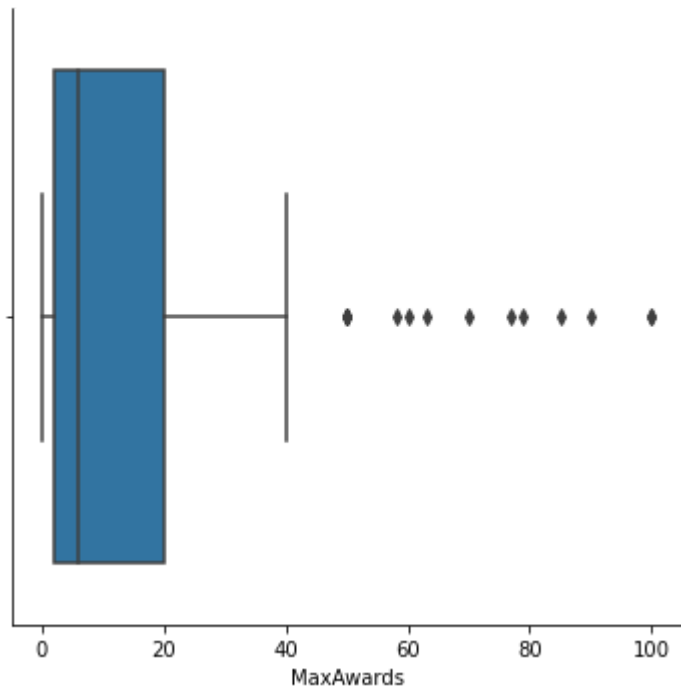
```
In [20]: sns.catplot(x = 'MaxAwards', kind = 'box', data = grants2) # still many outliers to pc
```

```
Out[20]: <seaborn.axisgrid.FacetGrid at 0x1d3930f9280>
```



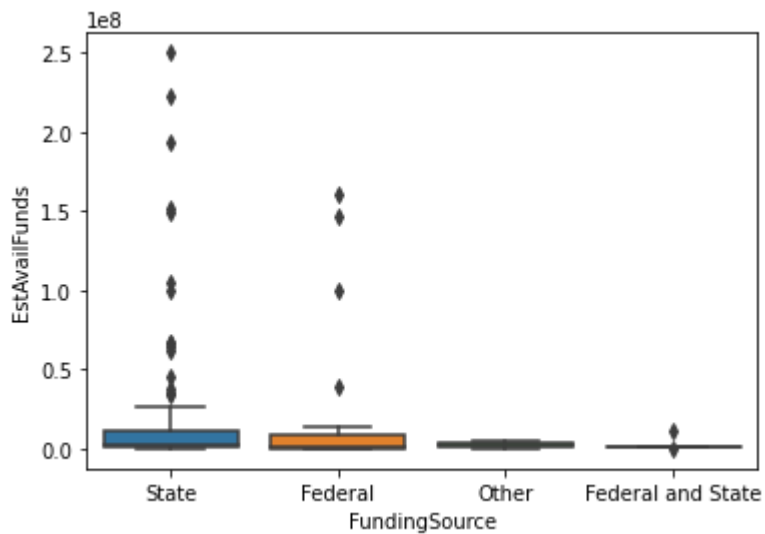
```
In [21]: grants3 = grants[grants["MaxAwards"] < 200] # further subset our data
sns.catplot(x = 'MaxAwards', kind = 'box', data = grants3)
```

```
Out[21]: <seaborn.axisgrid.FacetGrid at 0x1d393239670>
```



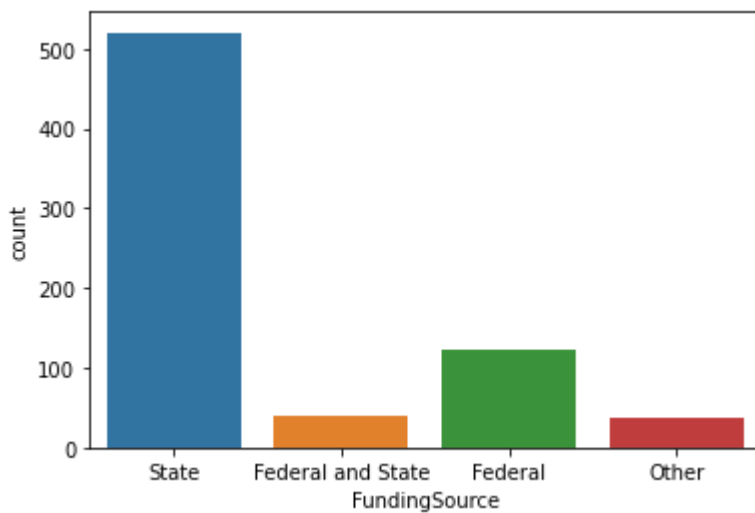
```
In [22]: # Potential relationship: Funding Source and Maximum Awards?
sns.boxplot(x = 'FundingSource', y = 'EstAvailFunds', data = grants2) #bulk of outlier
```

```
Out[22]: <AxesSubplot:xlabel='FundingSource', ylabel='EstAvailFunds'>
```



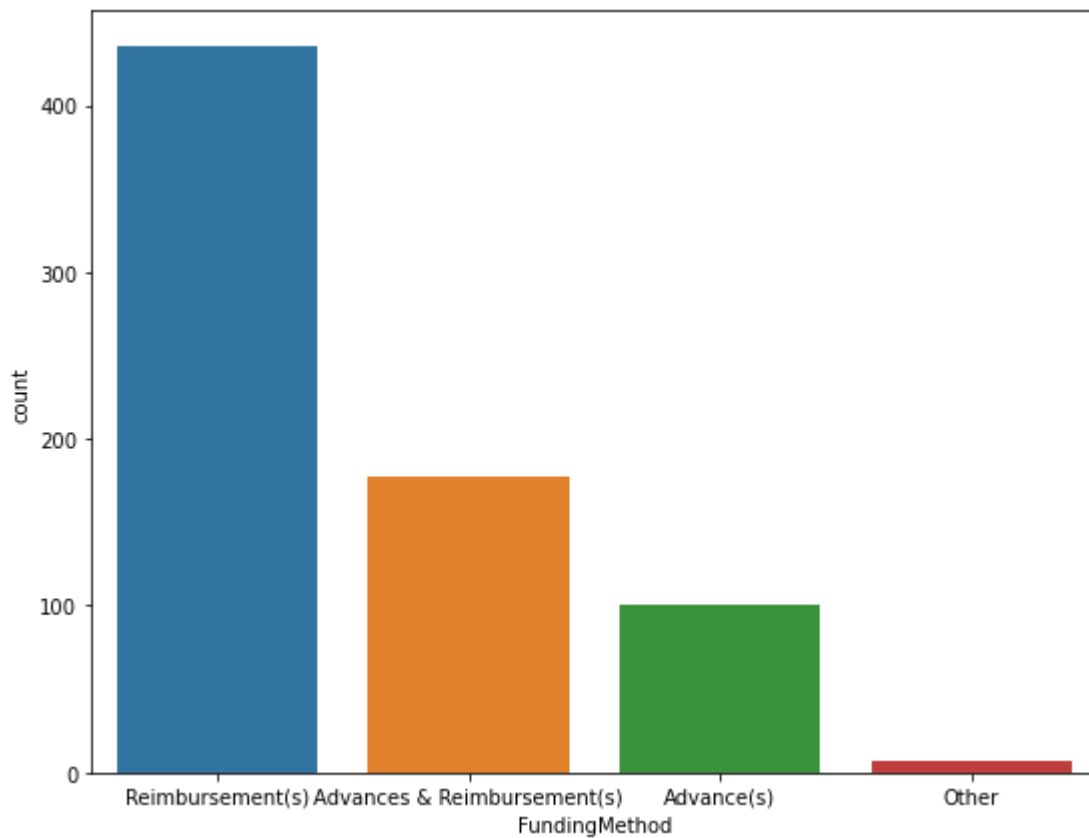
```
In [23]: sns.countplot(x = 'FundingSource', data = grants) #to be expected as we are dealing with
```

```
Out[23]: <AxesSubplot:xlabel='FundingSource', ylabel='count'>
```



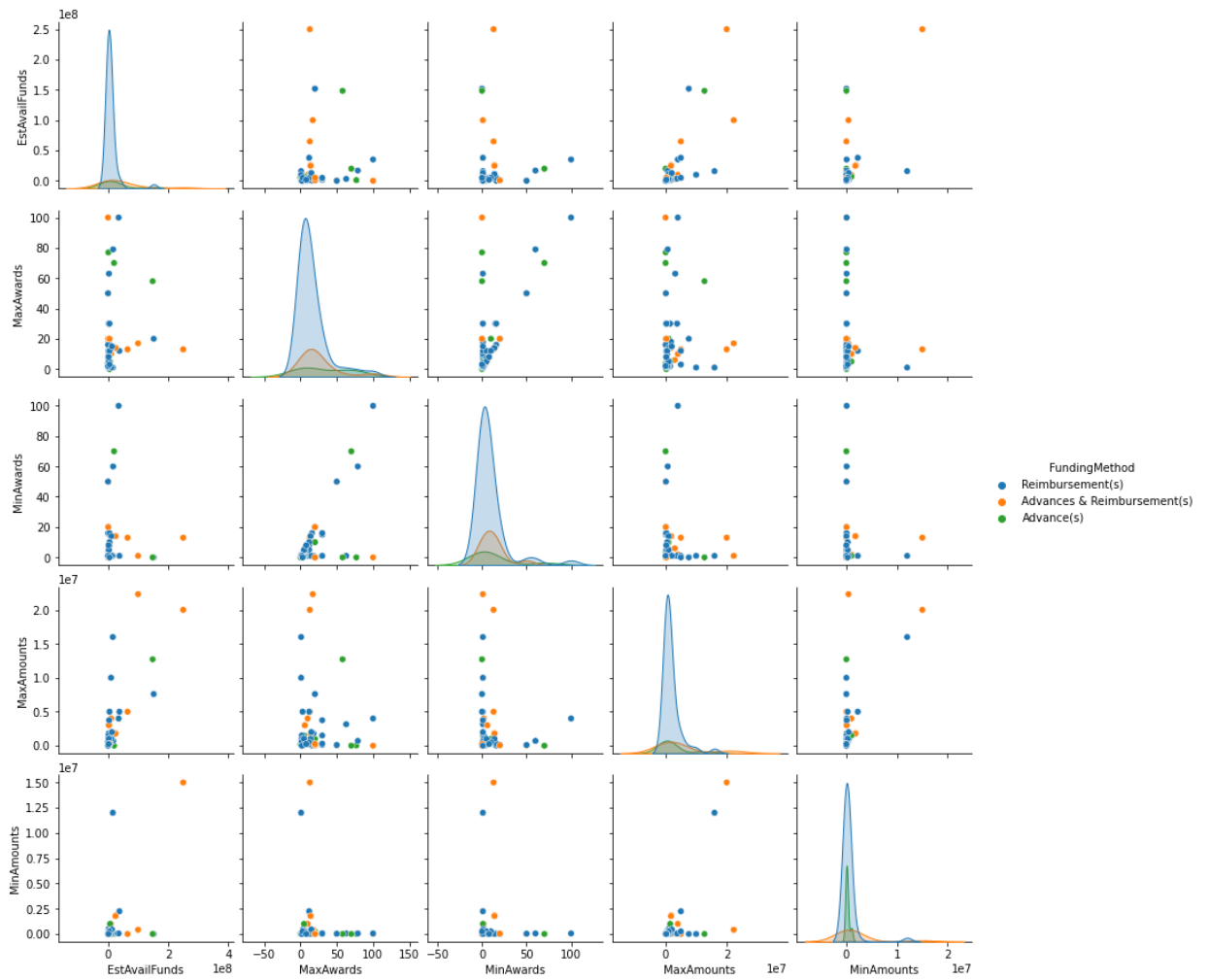
```
In [24]: # Another potentially interesting variable to consider: Funding Method
fig, ax = plt.subplots()
fig.set_size_inches(9,7)
sns.countplot(x = 'FundingMethod', data = grants, ax = ax)
```

```
Out[24]: <AxesSubplot:xlabel='FundingMethod', ylabel='count'>
```



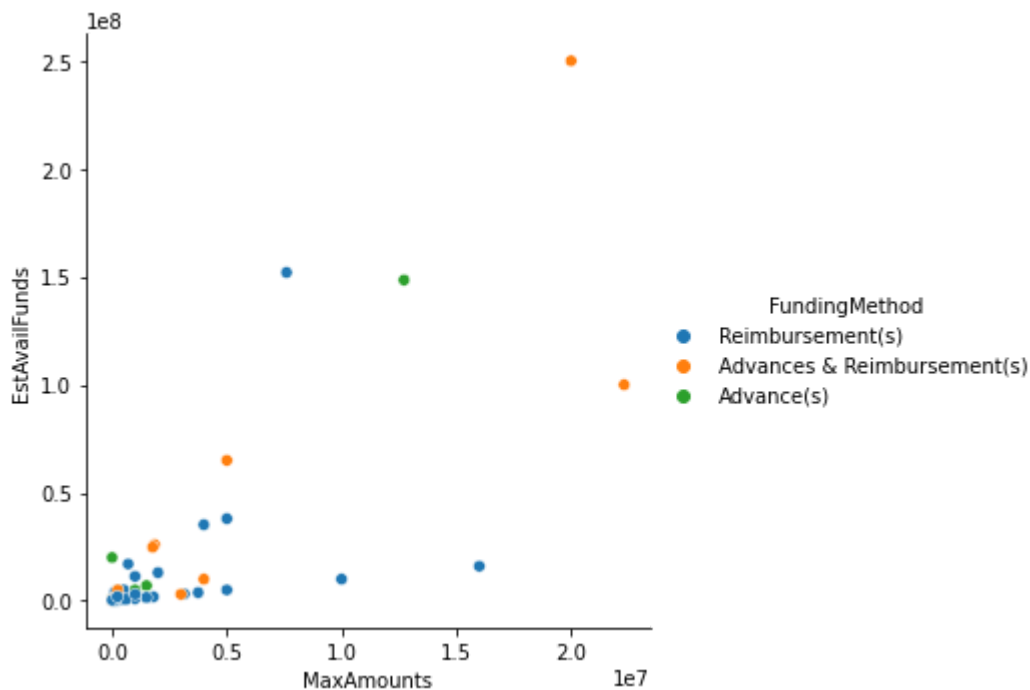
```
In [25]: # Looking further into funding method
grants4 = grants3.dropna(axis=0, how='any', thresh=None, subset=None, inplace=False) #
sns.pairplot(data = grants4.drop('PortalID', axis = 1), hue = 'FundingMethod')
```

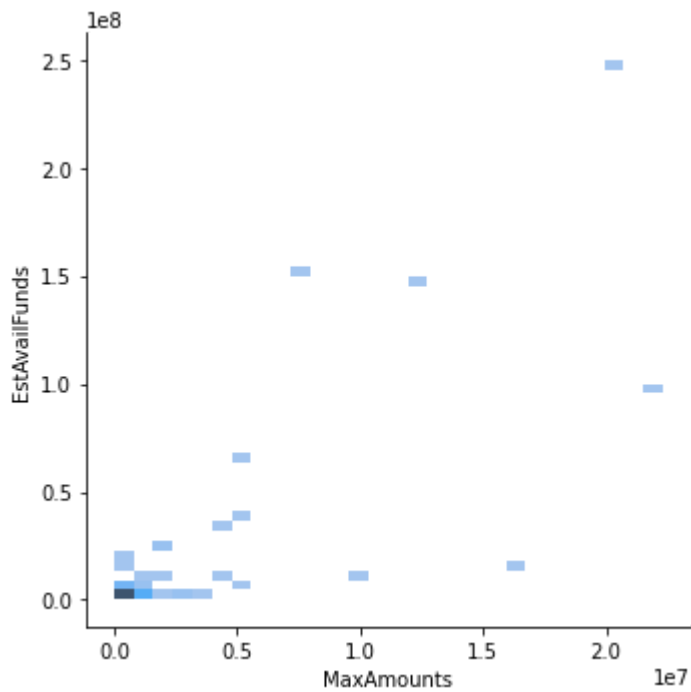
```
Out[25]: <seaborn.axisgrid.PairGrid at 0x1d394458250>
```



```
In [26]: # Most interesting scatter: Maximum Amount vs Estimated Available Funds?
sns.relplot(x = 'MaxAmounts', y = 'EstAvailFunds', hue = 'FundingMethod', data = grants)
sns.displot(data = grants4, x = 'MaxAmounts', y = 'EstAvailFunds')
```

Out[26]: <seaborn.axisgrid.FacetGrid at 0x1d3952b0790>





### Next Steps

- apply transformation
- linear model/analysis
- potential multivariate analysis as well?
- exploring the categories (split them up?)
- which agencies? whats useful?
- distribution of which agencies that are reporting stuff
- contextualize the data (are these values legit)
- are they even available?

## California Grants: Further Analysis

```
In [27]: grants.info()  
grants.ApplicationDeadline.value_counts().head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 725 entries, 0 to 724
Data columns (total 23 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   PortalID              725 non-null   object
1   Status                725 non-null   object
2   LastUpdated          725 non-null   object
3   AgencyDept           725 non-null   object
4   Title                725 non-null   object
5   Type                 725 non-null   object
6   LOI                  724 non-null   object
7   Categories           725 non-null   object
8   Purpose              724 non-null   object
9   Description           725 non-null   object
10  ApplicantType        721 non-null   object
11  FundingSource        721 non-null   object
12  MatchingFunds       725 non-null   object
13  EstAvailFunds       713 non-null   float64
14  FundingMethod        719 non-null   object
15  OpenDate             720 non-null   object
16  ApplicationDeadline  715 non-null   object
17  ExpAwardDate        713 non-null   object
18  GrantURL             725 non-null   object
19  MaxAwards            148 non-null   float64
20  MinAwards            148 non-null   float64
21  MaxAmounts          231 non-null   float64
22  MinAmounts          231 non-null   float64
dtypes: float64(5), object(18)
memory usage: 130.4+ KB
```

```
Out[27]: Ongoing          94
2021-10-29 17:00:00    9
2022-06-23 23:59:00    9
2021-02-23 17:00:00    8
2022-03-09 23:59:00    7
Name: ApplicationDeadline, dtype: int64
```

**While some grants do state if whether they are ongoing, some have dates in the near or far future**

- since this dataset and ongoing status is constantly changing find a way to determine the amount of objects that are ongoing
- a function that can be used for future iterations of this dataset?

```
In [28]: deadline = grants.ApplicationDeadline
ongoing = []
for i in deadline:
    if type(i) == float:
        ongoing.append(0)
    elif i[0] == '0':
        ongoing.append(1)
    elif i[0] == '2':
        temp = pd.to_datetime(i, format="%Y-%m-%d %H:%M:%S")
        today = pd.datetime.now()
        if temp < today:
            ongoing.append(0)
        else:
```



```
ongoing.append(1)
grants['IsOngoing'] = ongoing
```

```
C:\Users\joshu\AppData\Local\Temp\ipykernel_9288\4256576508.py:10: FutureWarning: The pandas.datetime class is deprecated and will be removed from pandas in a future version. Import from datetime module instead.
today = pd.datetime.now()
```

```
In [29]: grants.AgencyDept.unique()
```

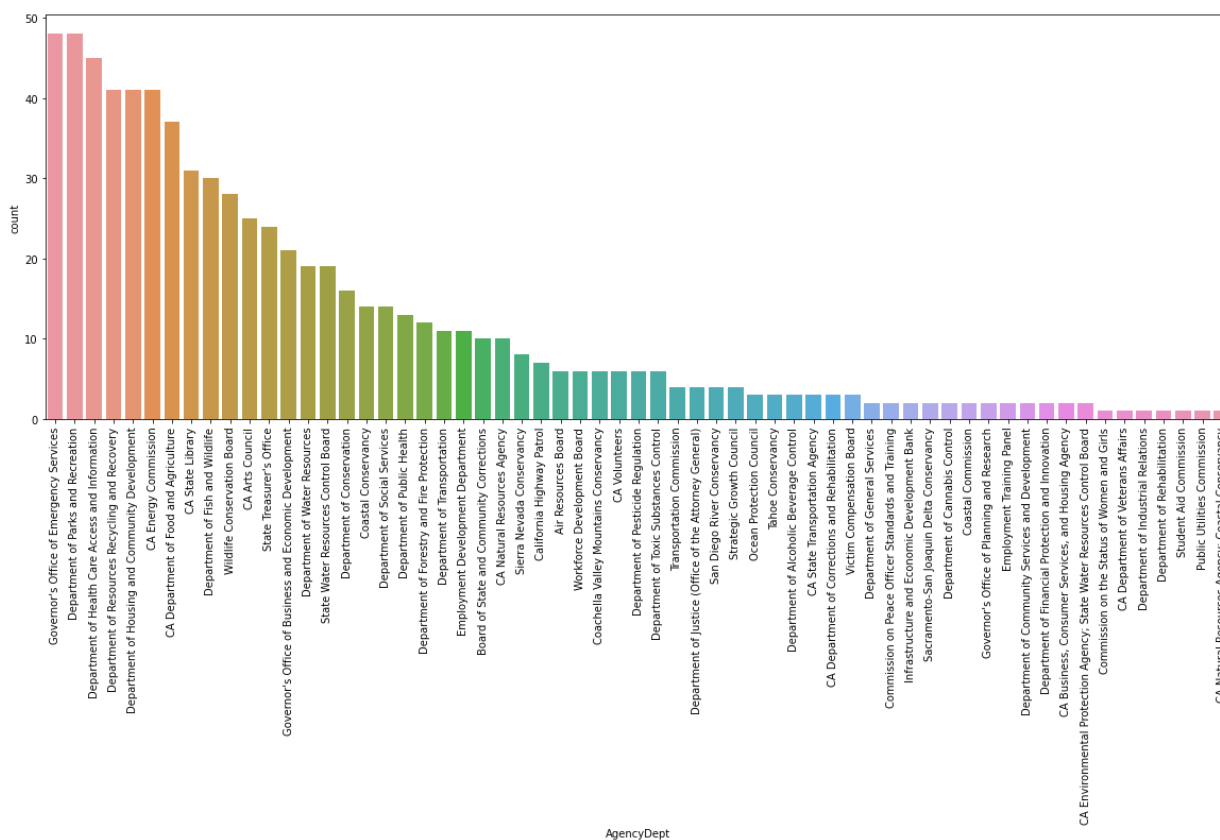
```
Out[29]: array(['Department of Justice (Office of the Attorney General)',
      'Department of Health Care Access and Information',
      'CA Arts Council', 'Department of Pesticide Regulation',
      "Governor's Office of Emergency Services",
      'Department of Social Services',
      'Department of Forestry and Fire Protection',
      'Board of State and Community Corrections', 'CA State Library',
      'Workforce Development Board',
      'Department of Parks and Recreation',
      "Governor's Office of Business and Economic Development",
      'Coachella Valley Mountains Conservancy',
      'Department of Conservation', 'Coastal Commission',
      'Department of Resources Recycling and Recovery',
      'CA Department of Food and Agriculture', 'CA Energy Commission',
      'CA Natural Resources Agency', 'Victim Compensation Board',
      'Department of Community Services and Development',
      'Department of Housing and Community Development',
      'Sierra Nevada Conservancy', 'Department of Water Resources',
      'Employment Development Department', 'Strategic Growth Council',
      'Wildlife Conservation Board', 'San Diego River Conservancy',
      'Department of Transportation', 'Ocean Protection Council',
      'Coastal Conservancy', 'Department of Fish and Wildlife',
      'State Water Resources Control Board', 'Air Resources Board',
      'Department of General Services', 'CA Volunteers',
      'Transportation Commission', 'Department of Public Health',
      'CA Business, Consumer Services, and Housing Agency',
      'Department of Financial Protection and Innovation',
      "State Treasurer's Office",
      'Department of Alcoholic Beverage Control',
      'Employment Training Panel', 'Department of Industrial Relations',
      "Governor's Office of Planning and Research",
      'Commission on the Status of Women and Girls',
      'California Highway Patrol',
      'Department of Toxic Substances Control',
      'CA Department of Corrections and Rehabilitation',
      'CA State Transportation Agency',
      'CA Department of Veterans Affairs',
      'Department of Cannabis Control', 'Department of Rehabilitation',
      'Sacramento-San Joaquin Delta Conservancy',
      'Student Aid Commission', 'Tahoe Conservancy',
      'Public Utilities Commission',
      'Infrastructure and Economic Development Bank',
      'Commission on Peace Officer Standards and Training',
      'CA Natural Resources Agency; Coastal Conservancy',
      'CA Environmental Protection Agency; State Water Resources Control Board'],
      dtype=object)
```

### Which Agencies are Reporting ?

```
In [30]: fig, ax = plt.subplots()
```

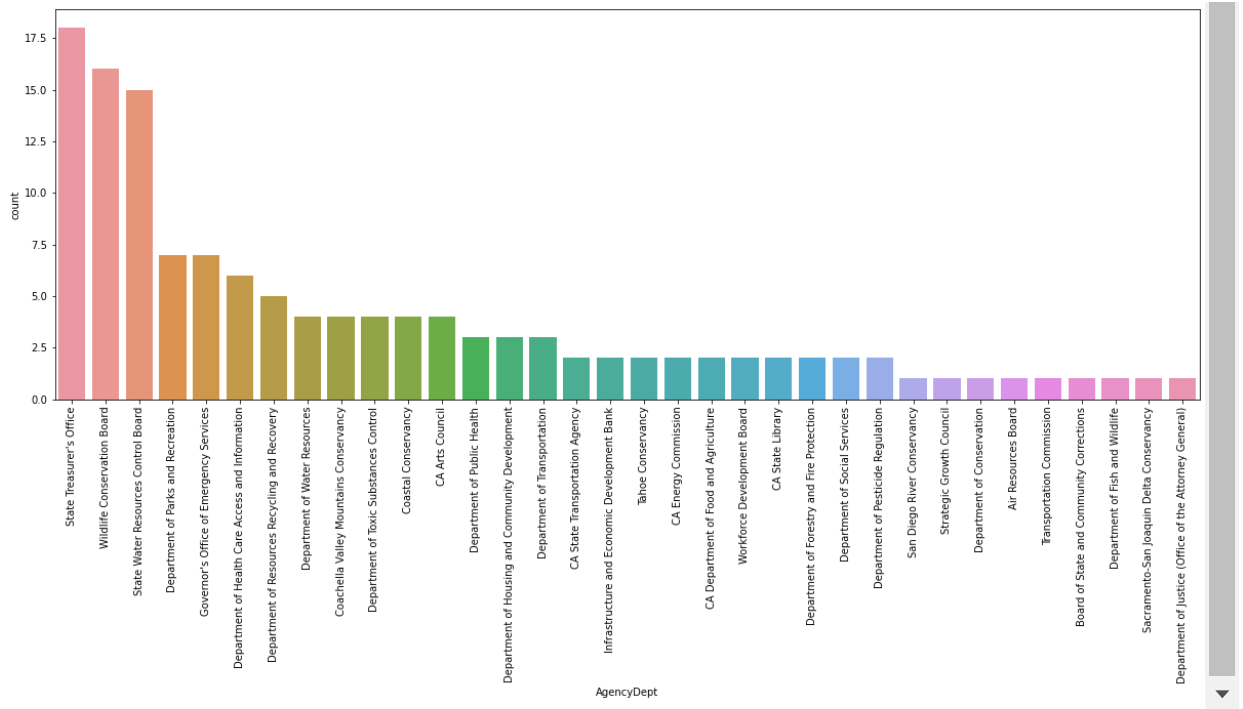
```
fig.set_size_inches(20,7)
plt.xticks(rotation = 90)
sns.countplot(x = 'AgencyDept', data = grants, ax = ax, order = grants.AgencyDept.value)
```

Out[30]: <AxesSubplot:xlabel='AgencyDept', ylabel='count'>



```
In [61]: # For ongoing grants?
grants_ongoing = grants[grants['IsOngoing'] == 1]
fig, ax = plt.subplots()
fig.set_size_inches(20,7)
plt.xticks(rotation = 90)
sns.countplot(x = 'AgencyDept', data = grants_ongoing, ax = ax, order = grants_ongoing
```

Out[61]: <AxesSubplot:xlabel='AgencyDept', ylabel='count'>



```
In [32]: grants.AgencyDept.unique() == grants_ongoing.AgencyDept.unique()
```

```
C:\Users\joshu\AppData\Local\Temp\ipykernel_9288\1530125029.py:1: DeprecationWarning:
elementwise comparison failed; this will raise an error in the future.
  grants.AgencyDept.unique() == grants_ongoing.AgencyDept.unique()
```

```
Out[32]: False
```

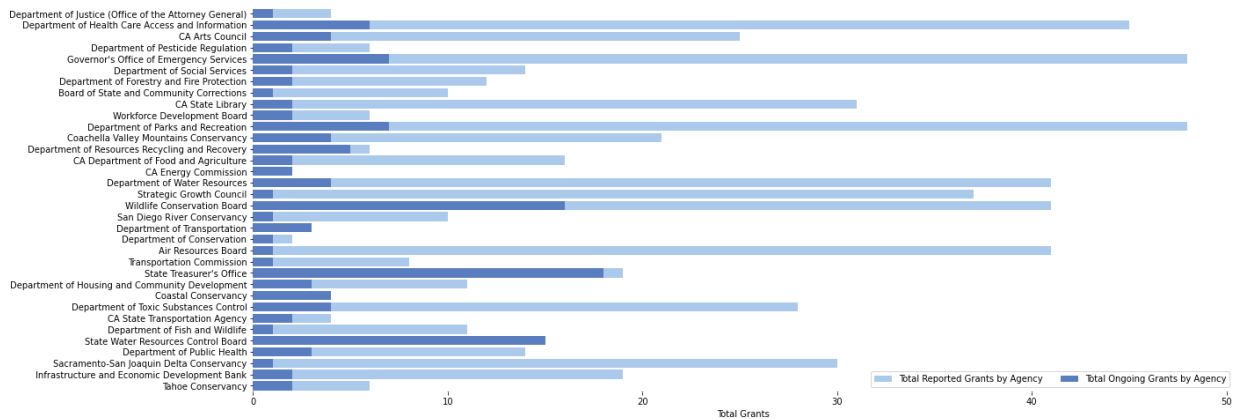
```
In [33]: # Initialize the matplotlib figure
f, ax = plt.subplots(figsize=(20,8))

# Plot
sns.set_color_codes("pastel")
sns.countplot(y = "AgencyDept", data = grants, label= "Total Reported Grants by Agency")

# Differentiate between ongoing grants
sns.set_color_codes("muted")
sns.countplot(y = "AgencyDept", data = grants_ongoing, label= "Total Ongoing Grants by")

# Add a Legend and informative axis Label
ax.legend(ncol=2, loc="lower right", frameon=True)
ax.set(ylabel="", xlabel="Total Grants")
sns.despine(left=True, bottom=True)

# Agencies with no ongoing grants are omitted
```



## Of ongoing grants, average amounts/awards by agency?

```
In [40]: print(grants_ongoing.MaxAwards.groupby(grants_ongoing.AgencyDept).mean().sort_values(ascending=True))
print(grants_ongoing.MaxAmounts.groupby(grants_ongoing.AgencyDept).mean().sort_values(ascending=True))
```

AgencyDept

Department of Pesticide Regulation	46.500000
State Water Resources Control Board	44.000000
Department of Water Resources	28.500000
State Treasurer's Office	20.000000
CA Arts Council	17.333333
Department of Parks and Recreation	11.333333
Department of Social Services	6.000000
Air Resources Board	1.000000
Board of State and Community Corrections	NaN
CA Department of Food and Agriculture	NaN

Name: MaxAwards, dtype: float64

AgencyDept

State Treasurer's Office	1.193111e+09
Department of Transportation	1.000000e+07
State Water Resources Control Board	2.941667e+06
Department of Pesticide Regulation	2.325000e+06
Strategic Growth Council	1.750000e+06
CA Department of Food and Agriculture	1.000000e+06
Department of Toxic Substances Control	9.833333e+05
Coachella Valley Mountains Conservancy	4.000000e+05
Board of State and Community Corrections	3.500000e+05
Department of Parks and Recreation	2.550000e+05

Name: MaxAmounts, dtype: float64

## Exploring the different categories of grants

```
In [36]: # create a potential function for a UI ?
grants_ongoing.Categories.unique()
```

```

Out[36]: array(['Education; Law, Justice, and Legal Services',
        'Health & Human Services',
        'Disadvantaged Communities; Libraries and Arts',
        'Disadvantaged Communities; Education; Employment, Labor & Training; Libraries
and Arts',
        'Agriculture; Disadvantaged Communities; Education; Employment, Labor & Traini
ng; Environment & Water; Food & Nutrition; Housing, Community and Economic Developmen
t; Parks & Recreation; Science, Technology, and Research & Development',
        'Law, Justice, and Legal Services',
        'Disadvantaged Communities; Education; Health & Human Services; Housing, Commu
nity and Economic Development',
        'Agriculture; Environment & Water',
        'Disadvantaged Communities; Education; Employment, Labor & Training; Law, Just
ice, and Legal Services',
        'Energy; Environment & Water; Food & Nutrition; Libraries and Arts; Science, T
echnology, and Research & Development',
        'Employment, Labor & Training',
        'Consumer Protection; Health & Human Services; Housing, Community and Economic
Development; Law, Justice, and Legal Services',
        'Education; Health & Human Services',
        'Consumer Protection; Disadvantaged Communities; Health & Human Services',
        'Environment & Water; Parks & Recreation', 'Environment & Water',
        'Agriculture; Disadvantaged Communities; Environment & Water',
        'Agriculture; Science, Technology, and Research & Development',
        'Environment & Water; Science, Technology, and Research & Development; Transpo
rtation',
        'Energy',
        'Disaster Prevention & Relief; Health & Human Services; Law, Justice, and Lega
l Services',
        'Consumer Protection; Health & Human Services; Law, Justice, and Legal Service
s',
        'Agriculture; Disadvantaged Communities; Employment, Labor & Training',
        'Consumer Protection; Disadvantaged Communities; Health & Human Services; Law,
Justice, and Legal Services',
        'Disadvantaged Communities; Environment & Water',
        'Disadvantaged Communities; Education',
        'Disadvantaged Communities',
        'Agriculture; Disadvantaged Communities; Disaster Prevention & Relief; Employm
ent, Labor & Training; Energy; Environment & Water; Housing, Community and Economic D
evelopment; Parks & Recreation',
        'Disadvantaged Communities; Environment & Water; Science, Technology, and Rese
arch & Development',
        'Environment & Water; Science, Technology, and Research & Development',
        'Disadvantaged Communities; Environment & Water; Science, Technology, and Rese
arch & Development; Transportation',
        'Disadvantaged Communities; Parks & Recreation',
        'Disadvantaged Communities; Environment & Water; Parks & Recreation',
        'Transportation', 'Environment & Water; Transportation',
        'Agriculture; Law, Justice, and Legal Services',
        'Disadvantaged Communities; Health & Human Services; Housing, Community and Ec
onomic Development',
        'Housing, Community and Economic Development',
        'Disaster Prevention & Relief; Environment & Water',
        'Disadvantaged Communities; Employment, Labor & Training; Energy; Science, Tec
hnology, and Research & Development; Transportation',
        'Environment & Water; Law, Justice, and Legal Services; Parks & Recreation',
        'Disadvantaged Communities; Disaster Prevention & Relief; Environment & Water;
Health & Human Services',
        'Disadvantaged Communities; Environment & Water; Housing, Community and Econom
ic Development',

```

```
'Disadvantaged Communities; Disaster Prevention & Relief; Environment & Water;
Science, Technology, and Research & Development',
'Disaster Prevention & Relief', 'Energy; Environment & Water',
'Disadvantaged Communities; Education; Housing, Community and Economic Develop
ment; Parks & Recreation',
'Education; Libraries and Arts',
'Energy; Environment & Water; Housing, Community and Economic Development; Lib
raries and Arts; Parks & Recreation',
'Disadvantaged Communities; Housing, Community and Economic Development',
'Agriculture; Disadvantaged Communities; Disaster Prevention & Relief; Environ
ment & Water; Housing, Community and Economic Development; Parks & Recreation',
'Education',
'Disaster Prevention & Relief; Health & Human Services',
'Disadvantaged Communities; Disaster Prevention & Relief; Employment, Labor &
Training; Energy; Environment & Water; Housing, Community and Economic Development; S
cience, Technology, and Research & Development; Transportation'],
dtype=object)
```

```
In [59]: # dictionary to tally each different category
cat = {}
for i in grants_ongoing.Categories:
    for j in i.split(';'):
        if j not in cat:
            cat[j] = 1
        else:
            cat[j] += 1

categories_data = pd.DataFrame({'category': cat.keys(), 'count': cat.values()})
categories_data
```

Out[59]:

	<b>category</b>	<b>count</b>
<b>0</b>	Education	9
<b>1</b>	Law, Justice, and Legal Services	8
<b>2</b>	Health & Human Services	8
<b>3</b>	Disadvantaged Communities	34
<b>4</b>	Libraries and Arts	5
<b>5</b>	Education	8
<b>6</b>	Employment, Labor & Training	8
<b>7</b>	Agriculture	11
<b>8</b>	Disadvantaged Communities	9
<b>9</b>	Environment & Water	34
<b>10</b>	Food & Nutrition	3
<b>11</b>	Housing, Community and Economic Development	14
<b>12</b>	Parks & Recreation	21
<b>13</b>	Science, Technology, and Research & Development	11
<b>14</b>	Law, Justice, and Legal Services	1
<b>15</b>	Health & Human Services	10
<b>16</b>	Energy	8
<b>17</b>	Employment, Labor & Training	3
<b>18</b>	Consumer Protection	4
<b>19</b>	Environment & Water	40
<b>20</b>	Transportation	6
<b>21</b>	Disaster Prevention & Relief	5
<b>22</b>	Disaster Prevention & Relief	5
<b>23</b>	Energy	3
<b>24</b>	Transportation	6
<b>25</b>	Housing, Community and Economic Development	3