HRS Data Exploration

Maria Navarro

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

pensions = pd.read_csv('/Users/marianavarro/Desktop/pdi_pensions/csv_files/h22j2_p.csv')
```

Notable Variables

SJ2Z503 - Job Type

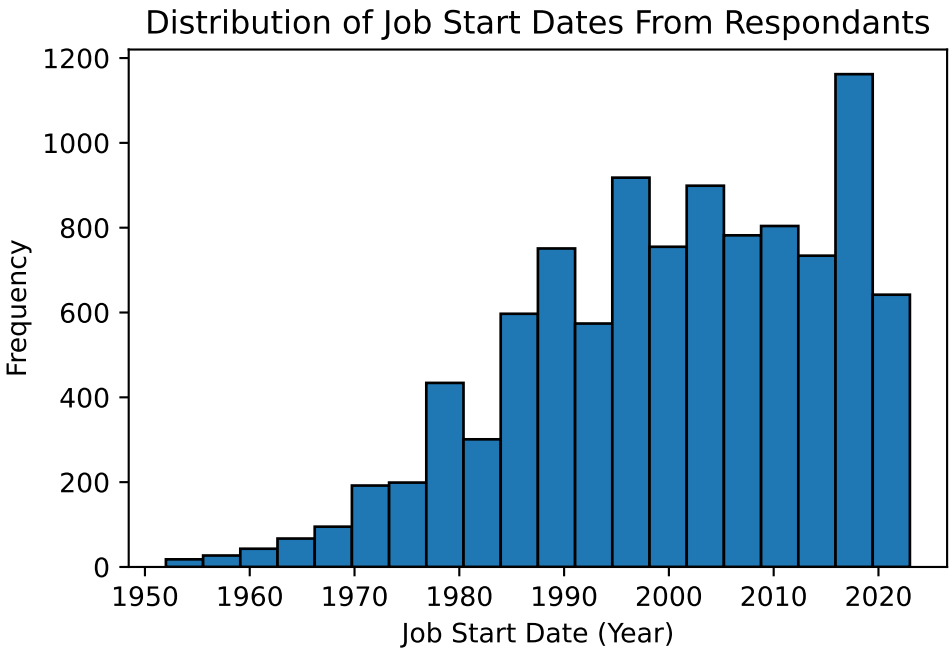
```
1.  DESIGNATED PREVIOUS WAVE JOB      3316
2.  OLD JOB                           3489
Blank.  Newly added past pension      3866
```

SJ2W410 - Year Job - Begin

```
This column will need to be cleaned as there are some job start years as low as -
8 and as high as 9999.
```

```
pensions['SJ2W410'].plot.hist(bins = 20, range = (1952, 2023), edgecolor = 'black')
plt.xlabel('Job Start Date (Year)')
plt.title('Distribution of Job Start Dates From Respondants')
print(pensions['SJ2W410'].max())
print(pensions['SJ2W410'].min())
```

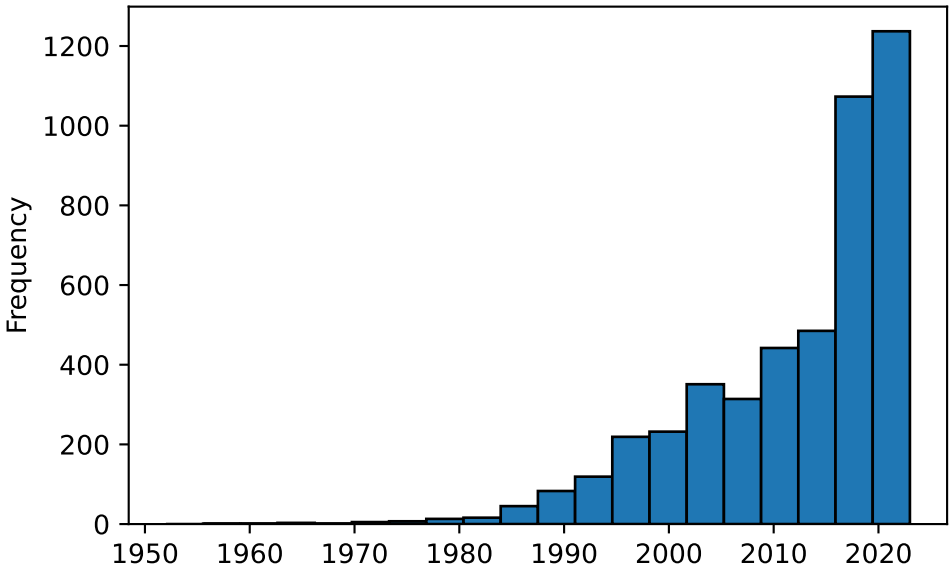
```
9999.0
-8.0
```



SJ2W411 - Year Job - End

This column will also need to be cleaned for same issue as Year Job - Begin.

```
pensions['SJ2W411'].plot.hist(bins = 20, range = (1952, 2023), edgecolor = 'black')
```



SJ2W405 - Plan Provider

		count
-8.	Web non-response	48
1.	PRIVATE EMPLOYER	5095
2.	GOVERNMENT EMPLOYER	2533
3.	R'S OWN BUSINESS	151
4.	UNION	495
97.	OTHER	430
98.	DK (Don't Know); NA (Not Ascertained)	89
99.	RF (Refused)	37
Blank.	INAP (Inapplicable); Partial Interview	1793