

Social Security Administration Analysis

Brielle Bernardy, Karen Fan, Haddi Fadia, and Jessica
Warner
Group 10





Introduction

Social Security benefits; funding that aids in retirement, disabled workers, and families where a parent/ spouse has passed

- June 2022- about 182 million people worked and paid Social Security taxes and about 66 million received Social Security benefits
 - Most beneficiaries are retirees and their families- 51 million people

How long will Social Security be around?

2022 annual report by the Social Security Board of Trustees shows that funding may be depleted by 2035.

Project Background

- What is the trend of social security payout over time (2010- 2020)?
- Average amount of payout per individual?
- Average amount of payout by State?
- Average amount of payout by Category (Retirement, disability, survivors)?

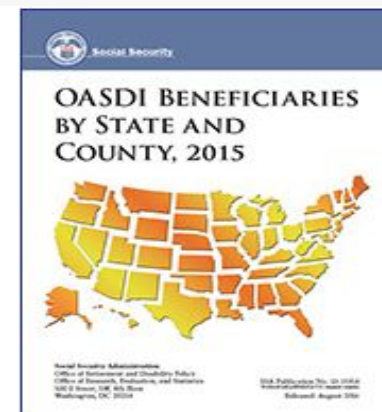


Data Collection

```
# Merged Files
df_2010 = pd.merge(aob_2010, nob_2010, on=["State or area", "County"], how='outer')
df_2015 = pd.merge(aob_2015, nob_2015, on=["State or area", "County"], how='outer')
df_2020 = pd.merge(aob_2020, nob_2020, on=["State or area", "County"], how='outer')
```

- Social Security Administration website
- Years: 2010, 2015, and 2020.
- The data is broken down by:
 - a. State or outlying areas,
 - b. County,
 - c. Type of benefit (retirement, survivors, disability), and
 - d. Sex of beneficiaries over age 65.
- The data was separated by the number of beneficiaries and amount of benefits paid.
- Saved csv files for our project, and merged the data together.
- Data is in thousands.

```
# Read CSV files
aob_2010 = pd.read_csv("data/aob_2010.csv", thousands=',')
aob_2015 = pd.read_csv("data/aob_2015.csv", thousands=',')
aob_2020 = pd.read_csv("data/aob_2020.csv", thousands=',')
nob_2010 = pd.read_csv("data/nob_2010.csv", thousands=',')
nob_2015 = pd.read_csv("data/nob_2015.csv", thousands=',')
nob_2020 = pd.read_csv("data/nob_2020.csv", thousands=',')
```



Social Security Administration. OASDI Beneficiaries by State and County, 2021. (n.d.). Retrieved February 2, 2023, from https://www.ssa.gov/policy/docs/statcomps/oasdi_sc/

Data Exploration and Cleanup

```
: # Convert data types for nob 2020
convert_dict = {"Total": float,
                "Retirement Retired workers": float,
                "Retirement Spouses": float,
                "Retirement Children": float,
                "Survivors Widow(er)s and parents": float,
                "Survivors Children": float,
                "Disability Disabled workers": float,
                "Disability Spouses": float,
                "Disability Children": float,
                "Aged 65 or older Men": float,
                "Aged 65 or older Women": float}

nob_2020 = nob_2020.astype(convert_dict)
```

Merged Files

```
df_2010 = pd.merge(aob_2010, nob_2010, on=["State or area", "County"], how='outer')
df_2015 = pd.merge(aob_2015, nob_2015, on=["State or area", "County"], how='outer')
df_2020 = pd.merge(aob_2020, nob_2020, on=["State or area", "County"], how='outer')
```

```
] # Sum of NaN values in df_2010
df_2010.isna().sum()
```

Drop outlying areas in df_2010

```
df_2010.drop(df_2010[df_2010['State or area'].str.contains('Outlying areas')].index, inplace=True)
```



**What is the trend of social
security payout over time?**



Correlation between the amount of benefits and the number of beneficiaries

```
In [31]: # Correlation between amount of benefits and number of beneficiaries 2010
correlation_df_2010 = pd.merge(states_people_2010, states_amount_2010, on="State or area", how="right")

# Correlation graph between amount of benefits and number of beneficiaries for 2010
people_2010 = correlation_df_2010.iloc[:,0]
amount_2010 = correlation_df_2010.iloc[:,1]
(slope, intercept, rvalue, pvalue, stderr) = linregress(people_2010, amount_2010)
regress_values = people_2010 * slope + intercept
line_eq = "y = " + str(round(slope,2)) + "x + " + str(round(intercept,2))
correlation = st.pearsonr(people_2010,amount_2010)
plt.figure(figsize=(8,5))
plt.scatter(people_2010,amount_2010)
plt.plot(people_2010, amount_2010, "r-")
plt.annotate(line_eq,(20,36),fontsize=15,color="red")
plt.title('2010 Amount of Benefits and the Number of Beneficiaries')
plt.xlabel('Total Amount by State (in thousands)')
plt.ylabel('Total Number of Beneficiaries by State (in thousands)')
print(f"The correlation between both factors is {round(correlation[0],2)}")

# Save the figure
plt.savefig("output_data/Correlation_2010.png")

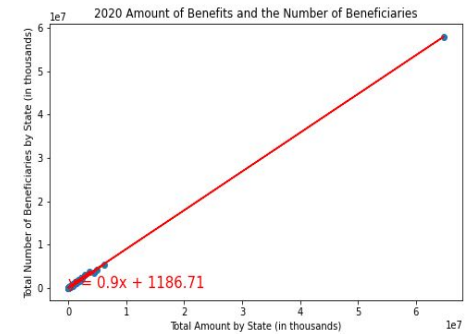
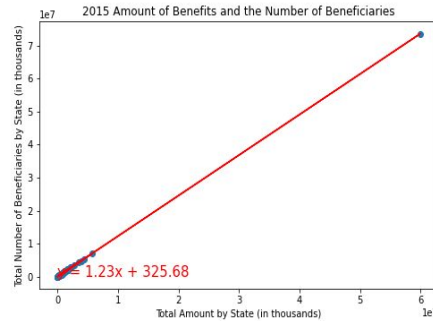
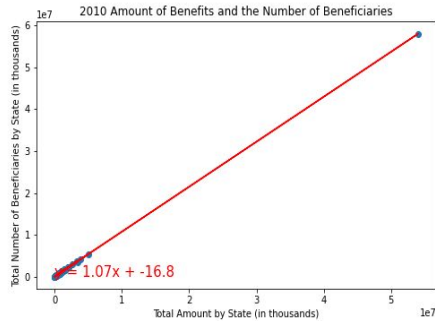
# Show graph
plt.show()
```

The correlation between both factors is 1.0

Correlation Between Amount of Benefits and Number of Beneficiaries

There is a pearson's r value of +1.0 for all three years sampled. This indicates a positive perfect linear correlation between the amount of benefits paid out verse the number of beneficiaries that were paid.

Incremental steps of one variable is exactly proportional to the value change of the other variable.



Average Payout Overtime

```
In [38]: # Average Payout Overtime
a_amount_2010 = df_2010['Total_amount'].sum()
a_people_2010 = df_2010['Total_people'].sum()
average_payout_2010 = a_amount_2010/a_people_2010
average_payout_2010 = average_payout_2010*1000

# Average Payout Overtime 2015
a_amount_2015 = df_2015['Total_amount'].sum()
a_people_2015 = df_2015['Total_people'].sum()
average_payout_2015 = a_amount_2015/a_people_2015
average_payout_2015 = average_payout_2015*1000

# Average Payout Overtime 2020
a_amount_2020 = df_2020['Total_amount'].sum()
a_people_2020 = df_2020['Total_people'].sum()
average_payout_2020 = a_amount_2020/a_people_2020
average_payout_2020 = average_payout_2020*1000
```

```
In [39]: # Average payout DataFrame
average_df = pd.DataFrame(
    {"Average Payout": [average_payout_2010, average_payout_2015, average_payout_2020]},
    index=['2010', '2015', '2020'])
average_df
```

Out[39]:

Average Payout	
2010	1074.663328
2015	1228.530419
2020	895.557122

Trend of Payouts over a Decade

Analysis

Average payouts for all states and outlying areas increased from 2010 through 2015 and then decreased from 2015 through 2020.

Minnesota and Wisconsin have similar patterns to the whole dataset.

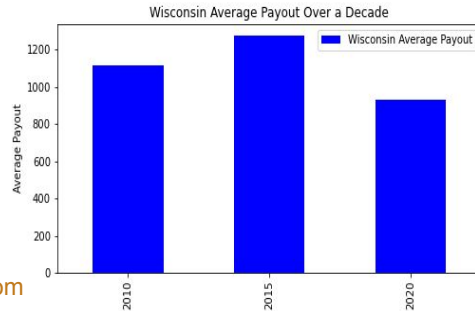
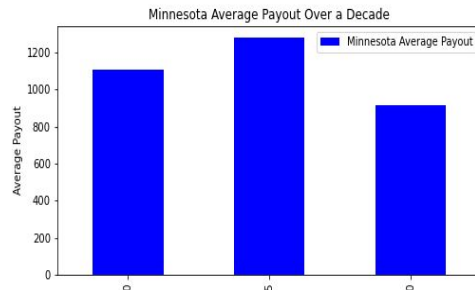
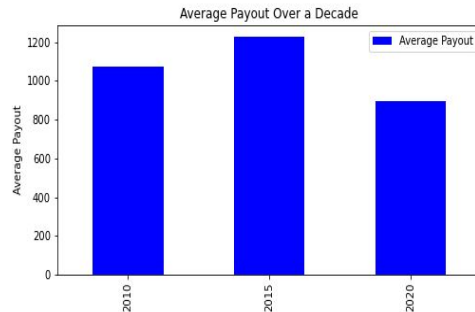
There was a 14.3% increase in the average payout from 2010 - 2015 in the month of December.

There was a 27.1% decrease in the average payout from 2015 - 2020 in the month of December.

If you look at the percent change over a 10 year period, there was a 16.6% decrease from 2010-2020 in average payout for the month of December.

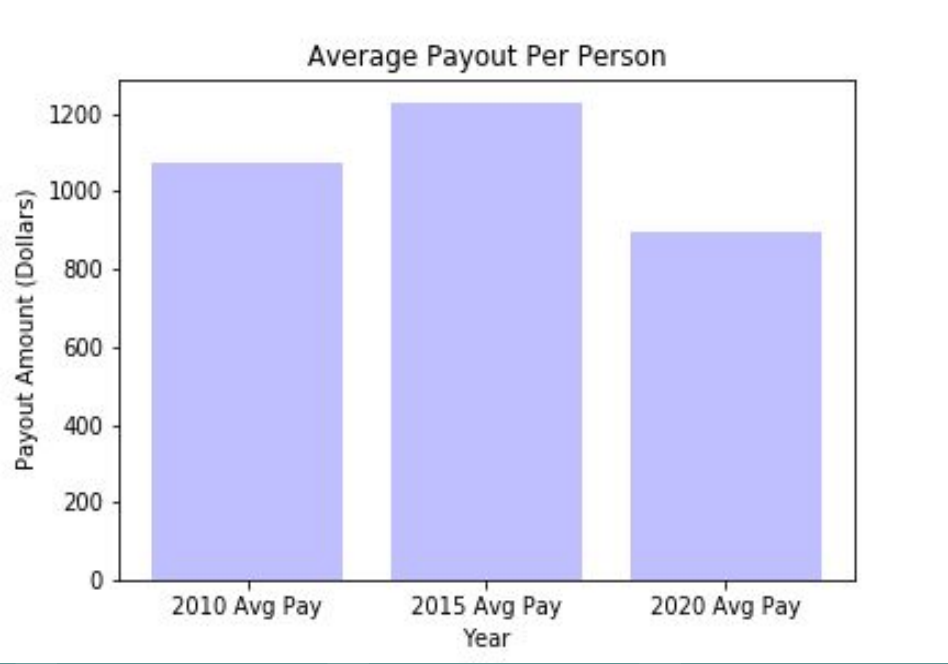
Follow-up Analysis

What caused the decrease in 2020? Was it due to the pandemic and elderly dying? Did less people working affect the amount of money brought in versus the amount of money being paid out?



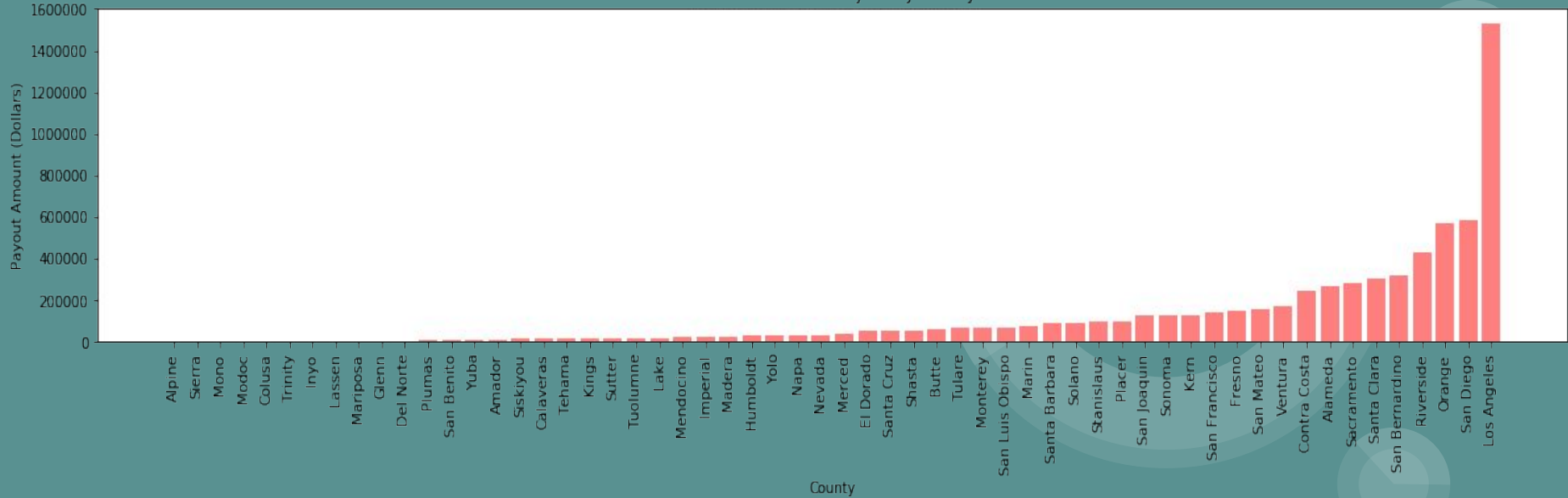
What is the average SS Payout per person?





	2010 Avg Pay	2015 Avg Pay	2020 Avg Pay
0	1074.332421	1228.115789	895.104378

California 2015 SS Payout by County



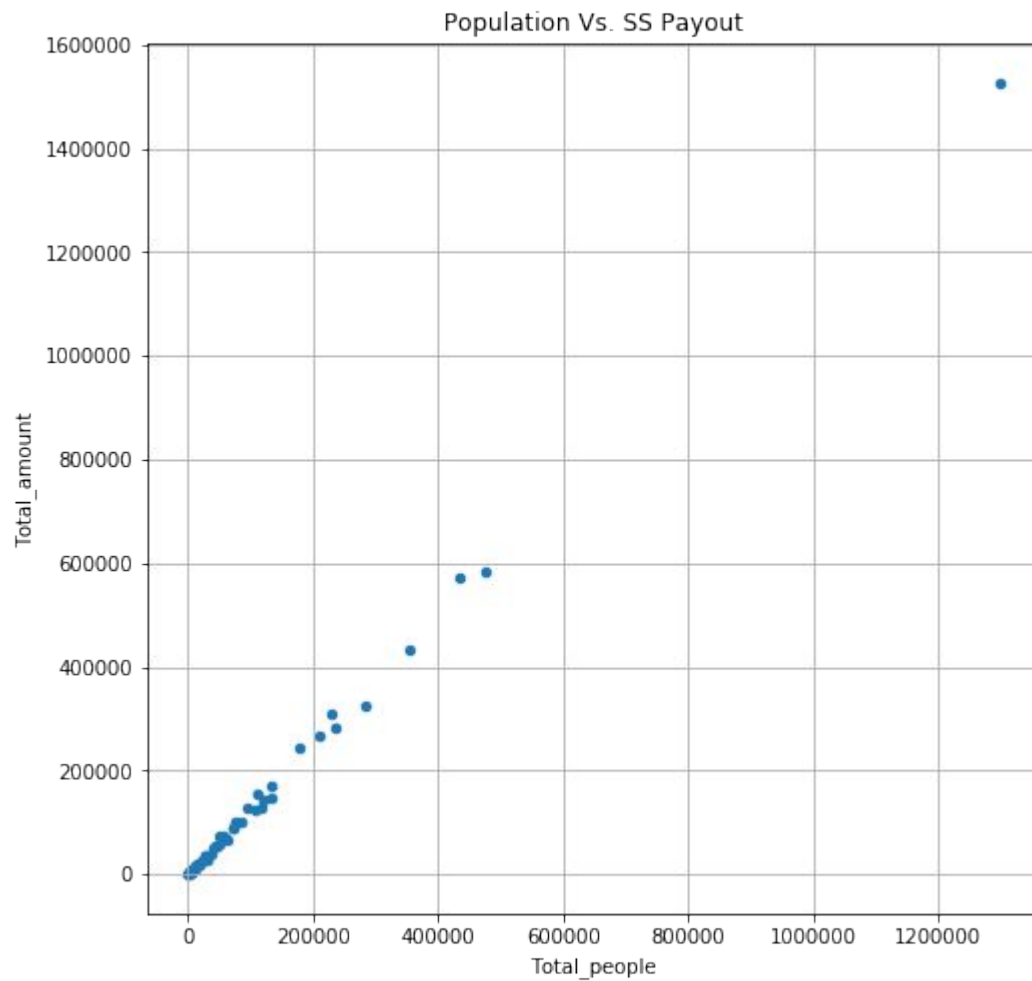
California receives the most SS in the United States

About 9% went to CA

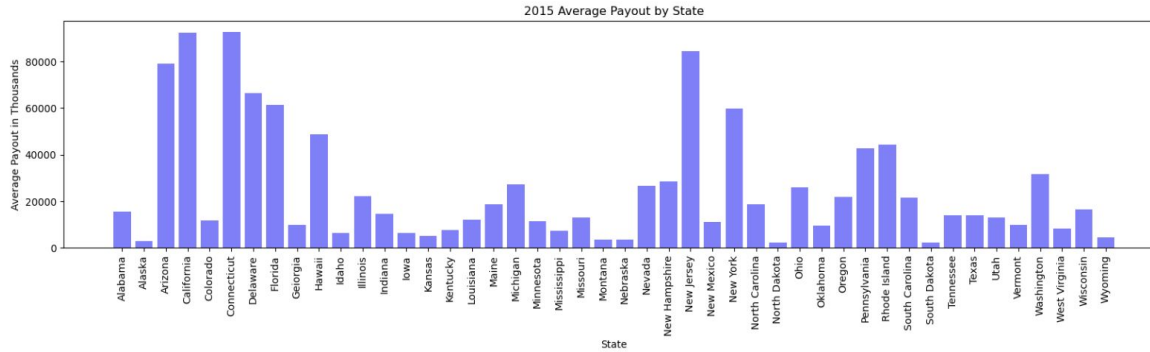
Most populous state in the US

Top 5 counties with the most payouts

LA most populous county with 9.9M people
22%

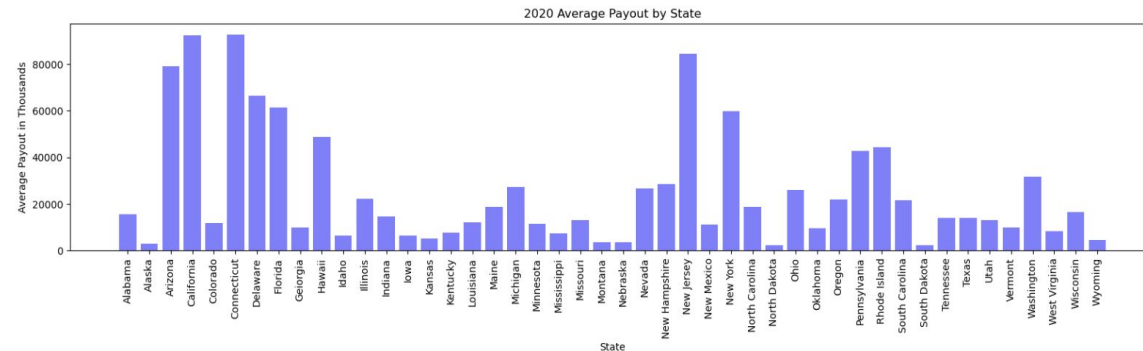


Average Payout by State - 2015 vs. 2020



Average Payout

count	48.000000
mean	31016.375000
std	31406.259639
min	2804.000000
25%	10291.500000
50%	17858.500000
75%	36419.000000
max	119014.000000



Average Payout

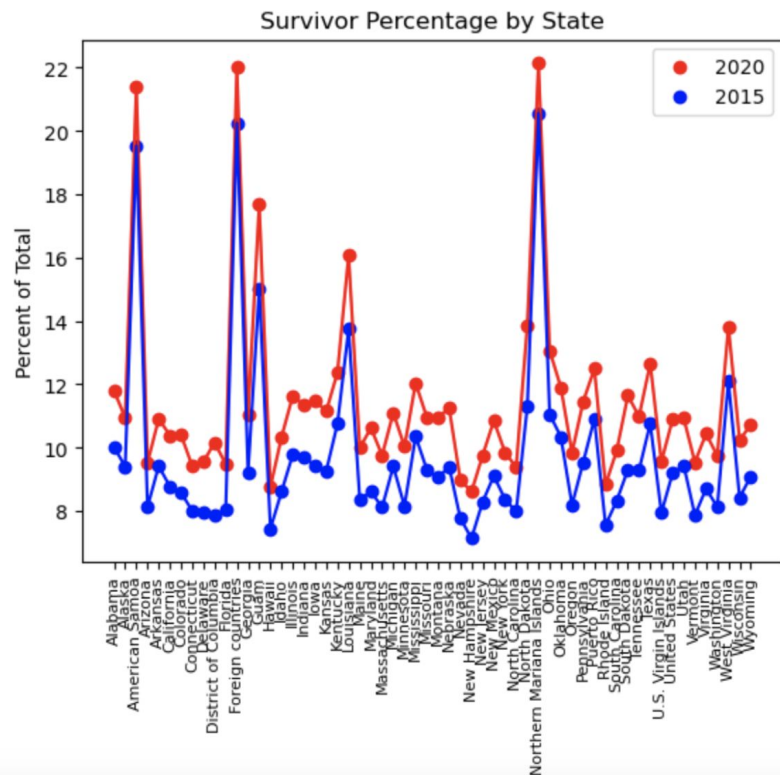
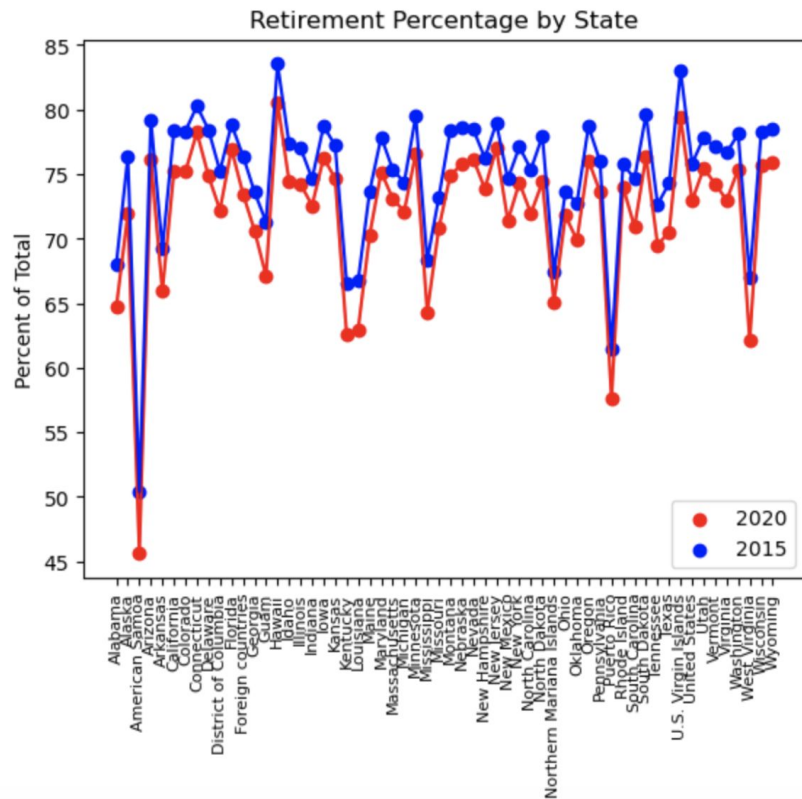
count	48.000000
mean	24374.166667
std	24773.692134
min	2297.000000
25%	8411.750000
50%	13903.000000
75%	27565.250000
max	92861.000000



Payout by Category: 2015 (top) vs 2020 (bottom)

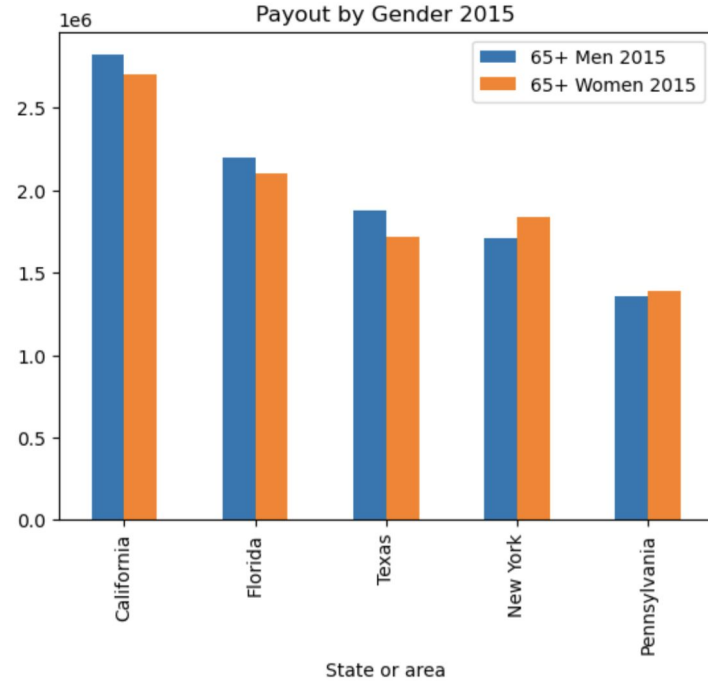
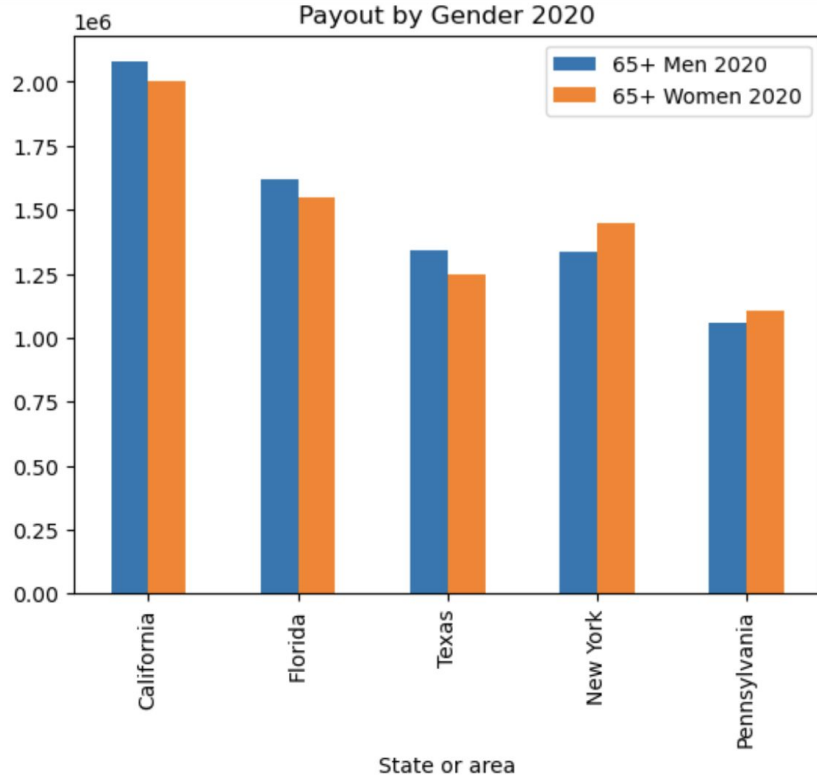
	Retirement Percentage	Survivor Percentage	Disability Percentage	65+ Percentage
State or area				
Alabama	68.031919	10.028244	21.938995	70.108615
Alaska	76.337454	9.396675	14.261240	75.658656
American Samoa	50.391886	19.525127	30.082988	51.175657
Arizona	79.205207	8.127840	12.667017	79.874034
Arkansas	69.200725	9.421195	21.377179	71.255117

	Retirement Percentage	Survivor Percentage	Disability Percentage	65+ Percentage
State or area				
Alabama	64.699863	11.802272	23.496997	66.131184
Alaska	71.894828	10.943972	17.141177	69.768606
American Samoa	45.596280	21.416849	32.986871	45.295405
Arizona	76.070787	9.523014	14.405608	75.138099
Arkansas	65.927923	10.910590	23.161329	67.424555



Category of interest: increase percentage in survivor category in 2020

Payout by Gender



The image features a solid orange background. In the top-left corner, there are three vertical bars of varying heights, each composed of three overlapping circles. In the bottom-right corner, there are four vertical bars of increasing height from left to right, each also composed of three overlapping circles.

Questions????