MEN VS. WOMEN

Who rules the sky (and beyond) when it comes to NASA hiring rates?

Kaity Tainer, Cynthia Cardenas, Elesha Hunter, Kendra Sawyer, Rachael Reich, Stacey Gonzalez

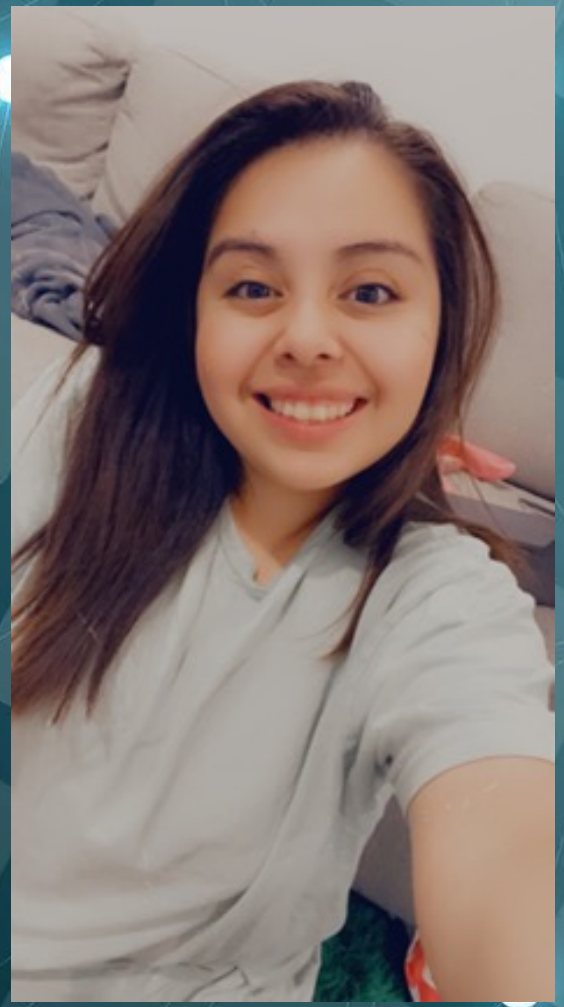
Rachael Reich

BS Mechanical Engineering (Drexel University, 2017)

Rachael is from NJ

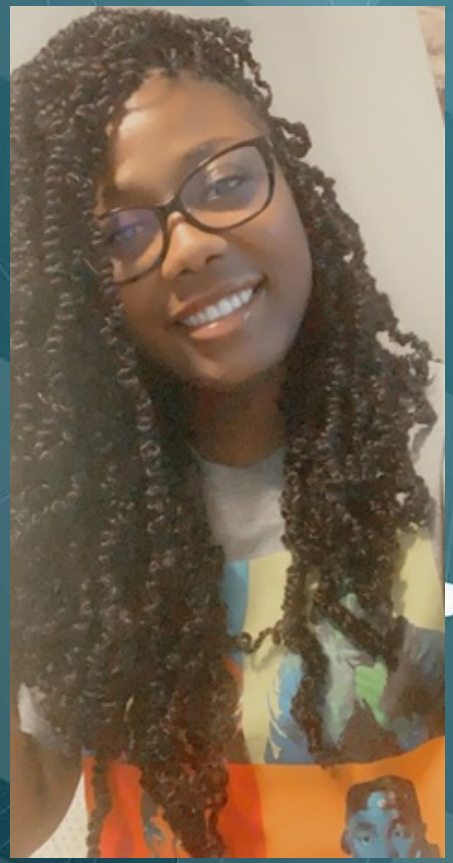
After three years in the Semiconductor industry, she transitioned to Data Science.

Cynthia Cardenas



BA Elementary Education (University of South Florida)
Currently employed: First-grade teacher

Elesha Hunter

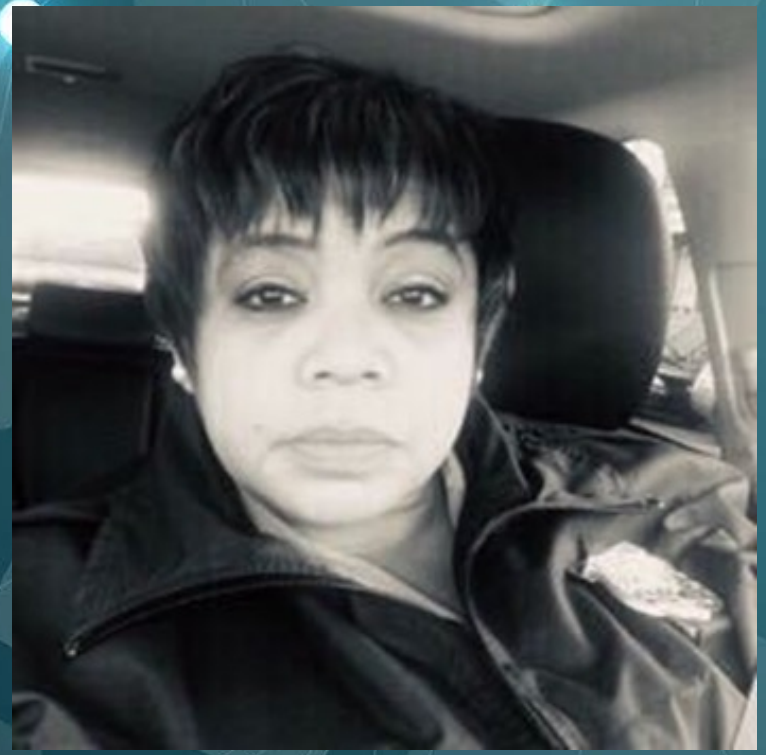


Elesha is from Michigan
Currently employed: Marketing Coordinator (San Antonio, TX)

Kendra Sawyer

BS Music Education (Elizabeth City State University)
Currently employed: Elementary Music Teacher

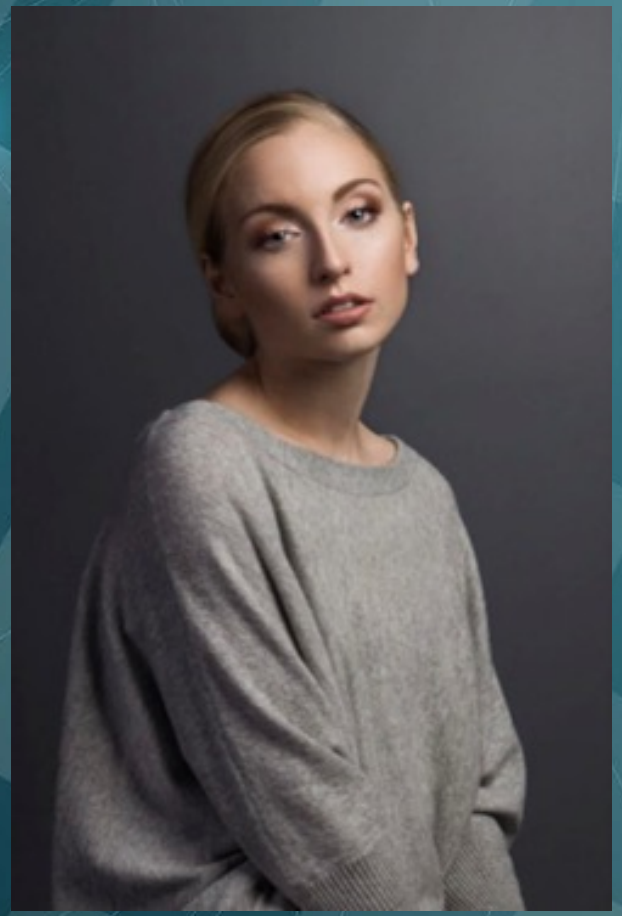
Stacey Gonzalez



BA Political Science
BS Sociology
MS Psychology MPH Public Health and Epidemiology
Phd International Psychology

Massive Incident Reaction Team certification from Homeland Security
Retired Commander of the United States Army Air Defense Artillery
Certified Officer of the Philadelphia Police Department and Critical Incident and Trauma certification.
NCIC / PCIC certification, SQL certification, Bias and Diversity certification from the Anti-Defamation League

Kaity Tainer



BA Psychology (University of Washington, 2016)

USPA Member (2016)

Currently Employed: Electrician (Tacoma, WA)
Commercial/Residential

Past employment: Executive Assistant (Los Angeles, CA)
Desilu-Studios (created Star Trek franchise) and Bruce Brown Films (The Endless Summer)

Background

- President Eisenhower suggested NASA chose from military test pilots
  - Also proposed NASA to congress
- 1959-1980: 65% came from the military
- Overall less women in the military
- Conversations about diversity in the workplace led us to more questions

Background

- The 70's: a springboard for women in astronomy
- 1977: recruitment of NASA skyrocketed because of Nichelle Nichols's help.

  Role as Lieutenant Uhura on Star Trek inspired young girls to become astronauts at NASA
- Also played a role recruiting people of color
- Received astronaut training and spoke to colleges about the importance of NASA
- Uhura translation (Swahili): "Freedom"

Methods

Data was gathered from the NASA website

•Data includes information on 714,193 astronauts. (sample size)

Wrangle Data

•We took multiple data sets and combined them in order to run our analysis

•We dropped unnecessary data

•Dropped missing data

•We recoded data from words (int64) to Boolean operators (True/False)

•Square root space walk and space flight (hr) in order for it to meet the normal distribution assumption for an independent t test

Variables

Age- mean: 38.158317, std: 7.615960, min: 24.5995,

max: 73.5633

- Education- mean- 7.548
- Space Flights- mean: 2.3599
- Space Flights (hr)- mean: 1210.822
- Space Walks- mean: 1.322
- Males Accepted in NASA Program- mean: 60.037
- Females Accepted in NASA Program- mean: 12.444
- Budget

Specific Methods Used:

- Python
- Independent t-test
- Correlation matrix
- Linear Regression
- Tableau

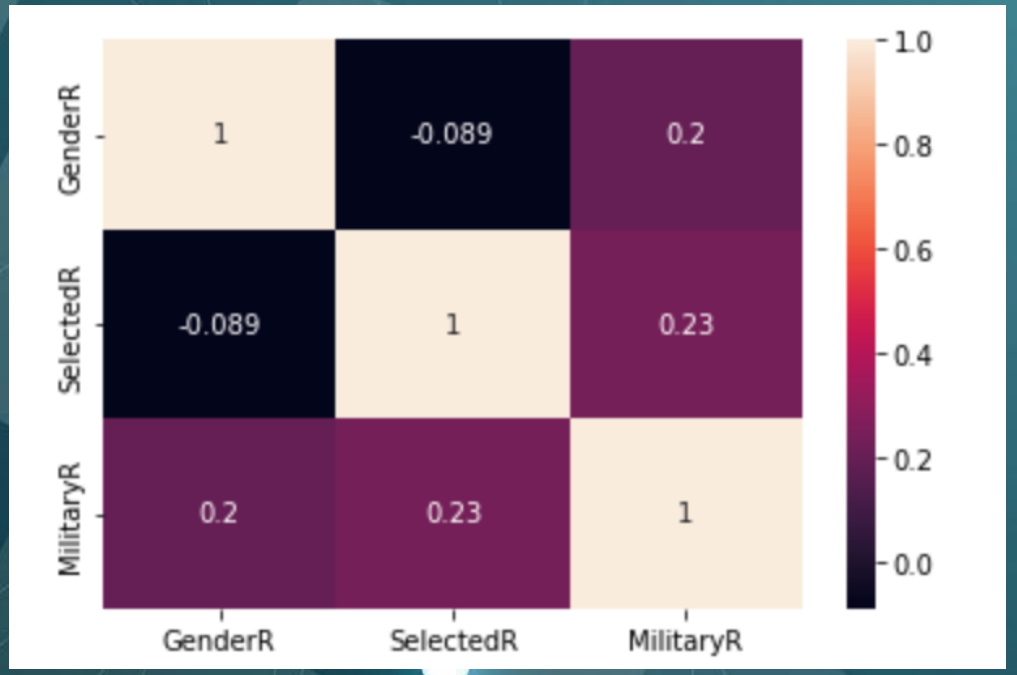
Evaluation Questions

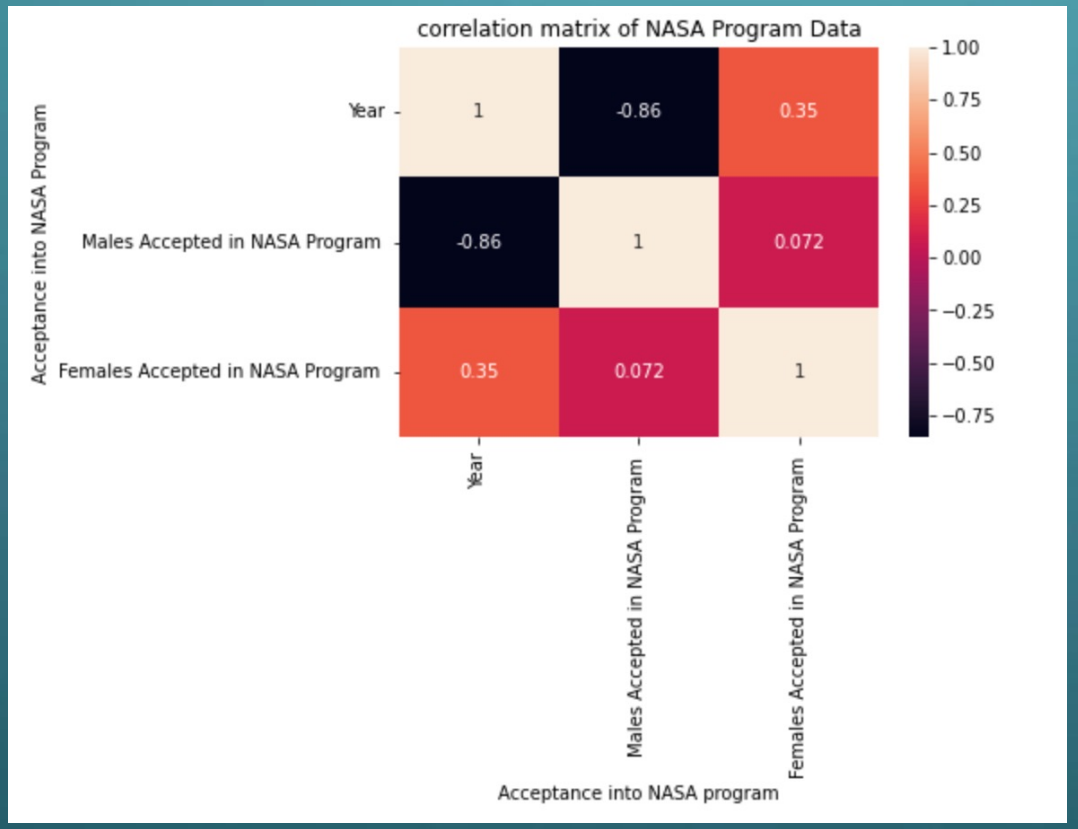
What is the difference between the hiring rate of males and females that are admitted into the NASA program in the last 10 years?

What is the ratio of males to females that are sent on space missions?

Correlations

- a=sns.heatmap(heatmap.corr(), annot=True)
- a=sns.heatmap(matrix.corr(), annot=True)





Simple Sample T Test

For males

```
stats.ttest_1samp(a=dataframe['Males Accepted in NASA Program '], popmean=50)
```

```
Ttest_1sampResult(statistic=2.7533640966067994, pvalue=0.010615850040028438)
```

- The null hypothesis, mean = 50
- Alternate hypothesis, the mean is not equal to 50
- Results:The p-value is less than 0.05, alternate hypothesis is correct.

For females

```
stats.ttest_1samp(a=dataframe['Females Accepted in NASA Program '], popmean=50)
```

```
Ttest_1sampResult(statistic=-13.716389697693476, pvalue=2.050532512478825e-13)
```

- The null hypothesis , mean = 50
- Alternate hypothesis , the mean is not equal to 50 results.
- The p-value is less than 0.05the alternate hypothesis is correct.

Independent T-test

```
#Independent T-Test
ttest_ind(dataframe['Females Accepted in NASA Program '], dataframe['Males Accepted in NASA Program '])
```

```
Ttest_indResult(statistic=-7.610115797517843, pvalue=5.281467527068827e-10)
```

- The null hypothesis , both the means are equal
- Alternate hypothesis , the means are not equal
- The p-value is less than 0.05, the alternate hypothesis is correct.

Linear Regression

Males by year

```
[ ]  #Linear Regression of Males by Year
     %matplotlib inline
     import  statsmodels.api as sm
     import statsmodels.stats.api as sms
     from scipy.stats import boxcox
     x = dataframe['Males Accepted in NASA Program ']
     y = dataframe['Year']
     model = sm.OLS(y,x).fit()
     model.summary()
```

OLS Regression Results

Dep. Variable:	Year	**R-squared (uncentered):**	0.786
Model:	OLS	**Adj. R-squared (uncentered):**	0.778
Method:	Least Squares	**F-statistic:**	95.45
Date:	Fri, 13 May 2022	**Prob (F-statistic):**	3.43e-10
Time:	23:02:33	**Log-Likelihood:**	-222.62
No. Observations:	27	**AIC:**	447.2
Df Residuals:	26	**BIC:**	448.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Males Accepted in NASA Program	22.7803	2.332	9.770	0.000	17.987	27.573

Omnibus:	6.777	**Durbin-Watson:**	0.098
Prob(Omnibus):	0.034	**Jarque-Bera (JB):**	6.147
Skew:	1.167	**Prob(JB):**	0.0463
Kurtosis:	2.864	**Cond. No.**	1.00

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

The R-square values shows that males count is dependent on year. Higher the R-square valur the higher the correlation.

Females by year

```
[ ]  # Linear Regression of Females by Year
     %matplotlib inline
     import  statsmodels.api as sm
     import statsmodels.stats.api as sms
     from scipy.stats import boxcox
     x = dataframe['Females Accepted in NASA Program ']
     y = dataframe['Year']
     model = sm.OLS(y,x).fit()
     model.summary()
```

OLS Regression Results

Dep. Variable:	Year	**R-squared (uncentered):**	0.446
Model:	OLS	**Adj. R-squared (uncentered):**	0.425
Method:	Least Squares	**F-statistic:**	20.95
Date:	Fri, 13 May 2022	**Prob (F-statistic):**	0.000103
Time:	23:06:12	**Log-Likelihood:**	-235.45
No. Observations:	27	**AIC:**	472.9
Df Residuals:	26	**BIC:**	474.2
Df Model:	1		
Covariance Type:	nonrobust		

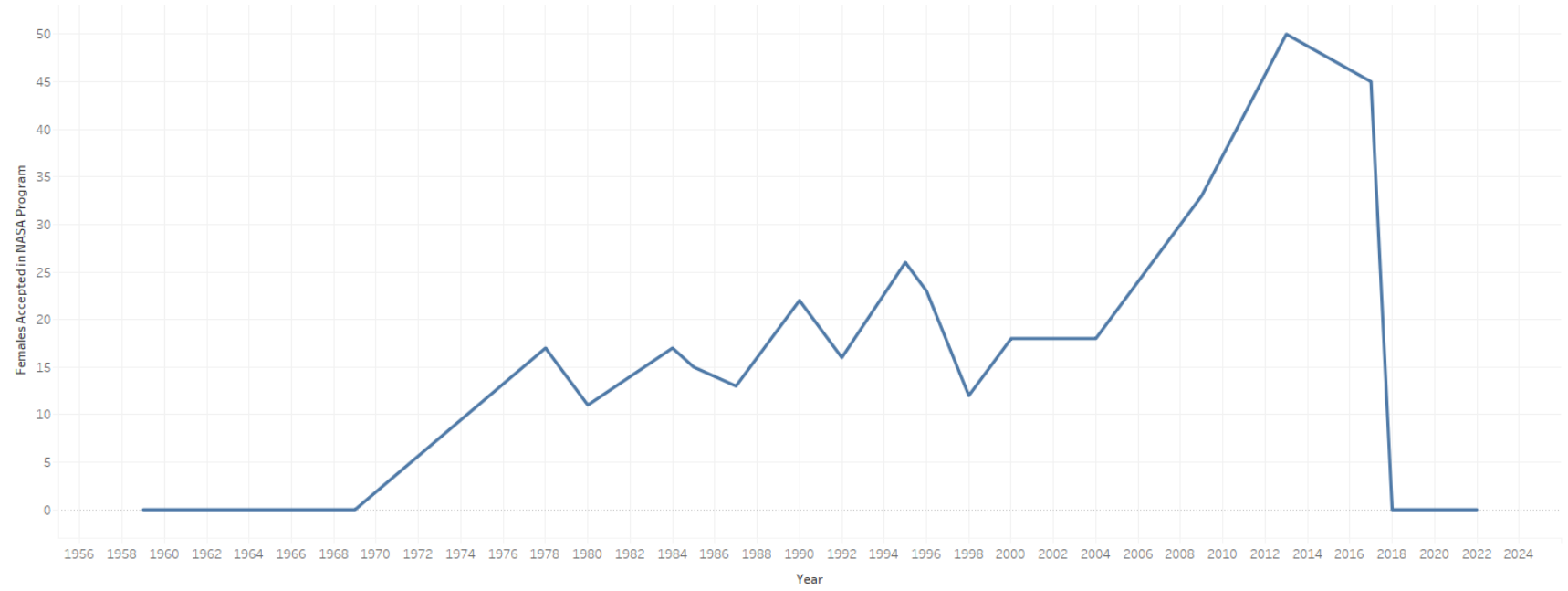
	coef	std err	t	P>|t|	[0.025	0.975]
Females Accepted in NASA Program	71.1489	15.543	4.577	0.000	39.199	103.099

Omnibus:	6.902	**Durbin-Watson:**	0.282
Prob(Omnibus):	0.032	**Jarque-Bera (JB):**	5.291
Skew:	-1.052	**Prob(JB):**	0.0710
Kurtosis:	3.522	**Cond. No.**	1.00

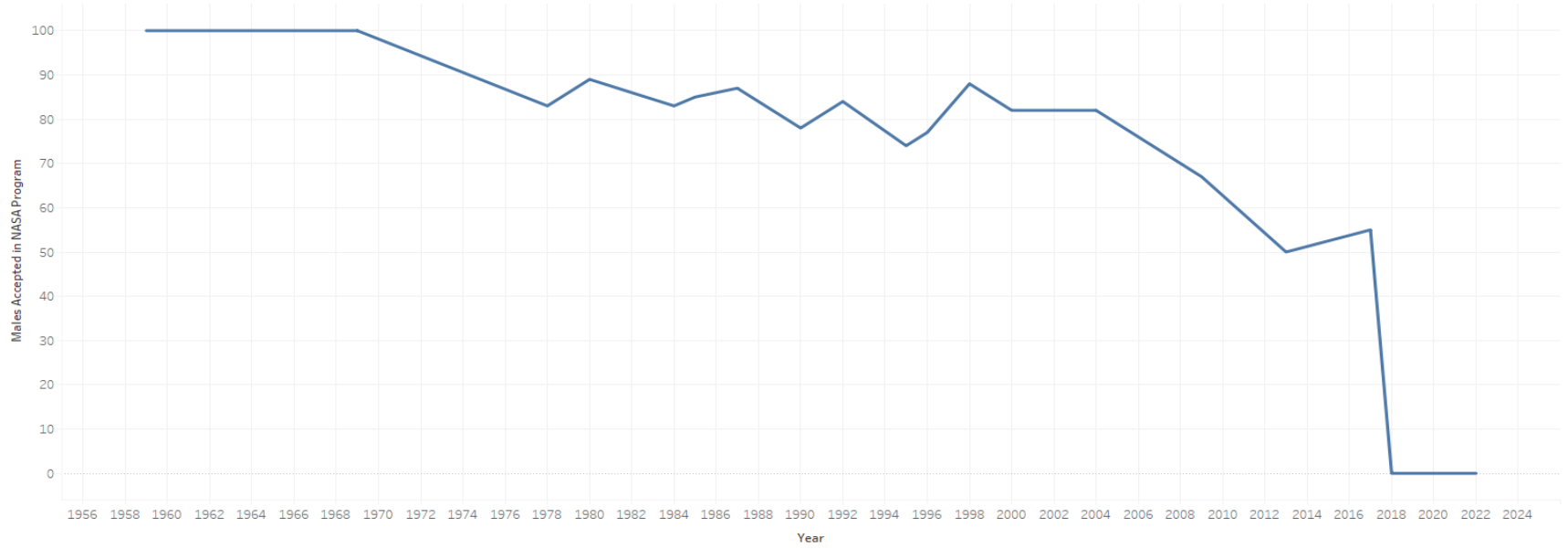
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

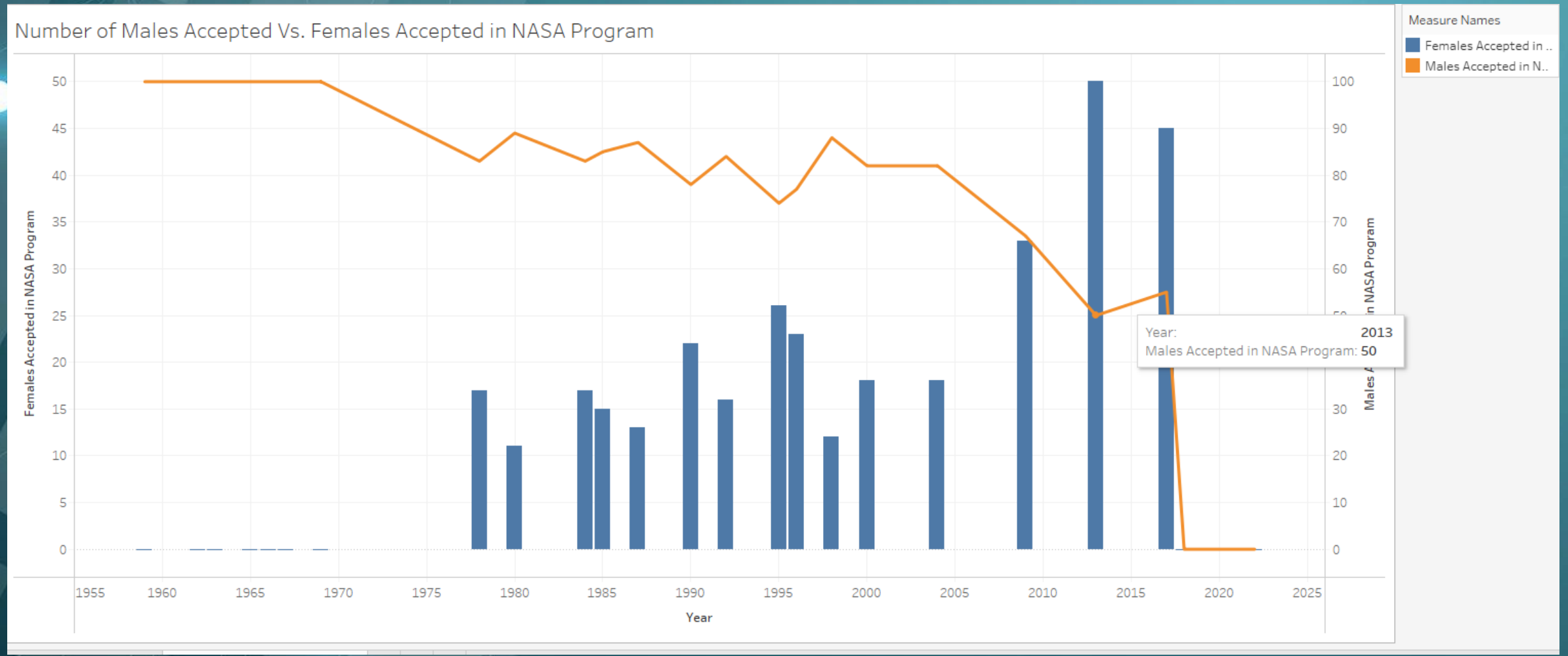
The R-square shows that the year and females are not correlated; as the value is 0.4 which is far from 1, which is a weak correlation .

Females Accepted in NASA Program



Males Accepted in NASA Program

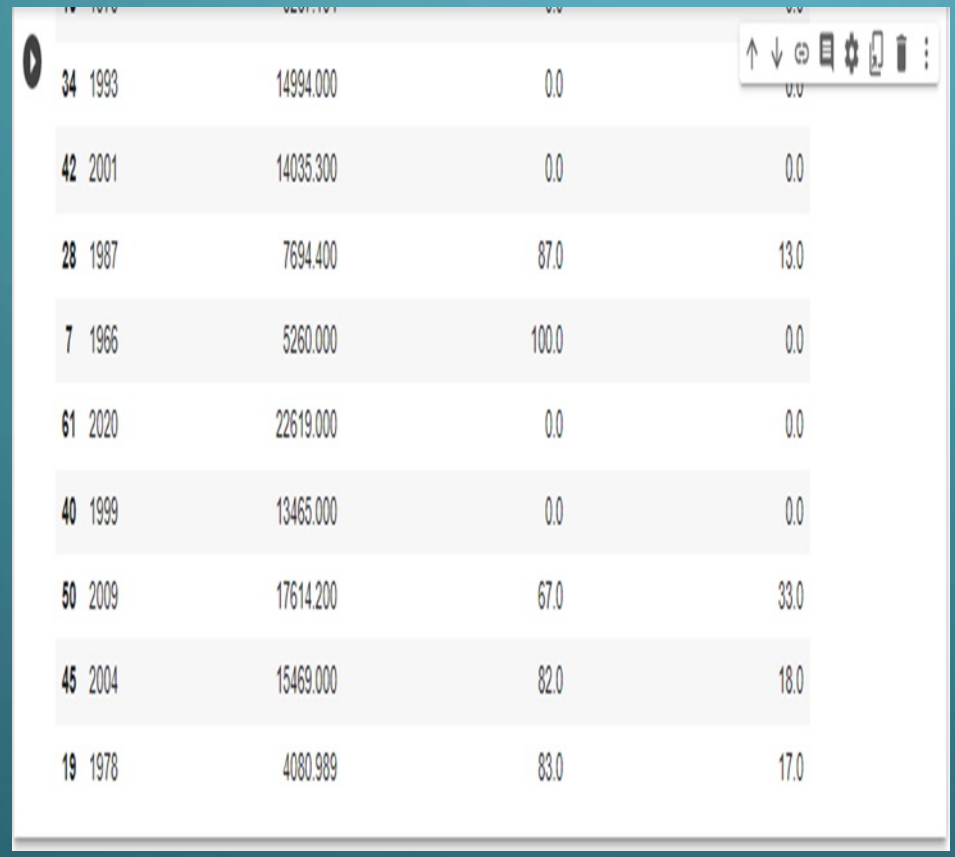



Number of Males Accepted Vs. Females Accepted in NASA Program
Measure Names
Females Accepted in ..
Males Accepted in N..
Females Accepted in NASA Program
Males Accepted in NASA Program
Year
Year: 2013
Males Accepted in NASA Program: 50
0
5
10
15
20
25
30
35
40
45
50
0
10
20
30
60
70
80
90
100
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025

Data Frame of Budget Analysis

```
#1. What is the difference between men getting accepted versus women?
dataframe.sample(n=30,random_state=42)
```

	Year	White House Budget Submission	Males Accepted in NASA Program	Females Accepted in NASA Program
22	1981	5736.654	0.0	0.0
0	1959	426.600	100.0	0.0
47	2006	16456.300	0.0	0.0
4	1963	3787.300	100.0	0.0
53	2012	18724.300	0.0	0.0
18	1977	3728.777	0.0	0.0
10	1969	4370.400	100.0	0.0
33	1992	15722.694	84.0	16.0
44	2003	15000.000	0.0	0.0
12	1971	3376.944	0.0	0.0
31	1990	13273.995	78.0	22.0
9	1968	5100.000	0.0	0.0
59	2018	19092.000	0.0	0.0
5	1964	5712.000	0.0	0.0
68	2027	0.000	0.0	0.0
30	1989	11488.000	0.0	0.0
57	2016	18529.100	0.0	0.0
35	1994	15266.000	0.0	0.0
56	2015	17460.000	0.0	0.0
46	2005	16224.000	0.0	0.0
16	1975	3267.104	0.0	0.0
34	1993	14994.000	0.0	0.0



34	1993	14994.000	0.0	0.0
42	2001	14035.300	0.0	0.0
28	1987	7694.400	87.0	13.0
7	1966	5260.000	100.0	0.0
61	2020	22619.000	0.0	0.0
40	1999	13465.000	0.0	0.0
50	2009	17614.200	67.0	33.0
45	2004	15469.000	82.0	18.0
19	1978	4080.989	83.0	17.0

Single Sample T-test for Females vs Male Budget Analysis

```
[ ]  # Single Sample t-test for Females
     import scipy.stats as stats
     #perform one sample t-test
     stats.ttest_1samp(a=dataframe['Females Accepted in NASA Program '], popmean=30)
```

```
Ttest_1sampResult(statistic=-19.481120762275467, pvalue=1.5727116412325747e-29)
```

1. The null hypothesis , mean = 30
2. Alternaye hypothesis , the mean is not equal to 30 results. The p-value is less than 0.05 and the null hypothesis is rejected and the alternate hypothesis is correct.

```
[ ]  # Single Sample t-test for Males
     import scipy.stats as stats
     #perform one sample t-test
     stats.ttest_1samp(a=dataframe['Males Accepted in NASA Program '], popmean=50)
```

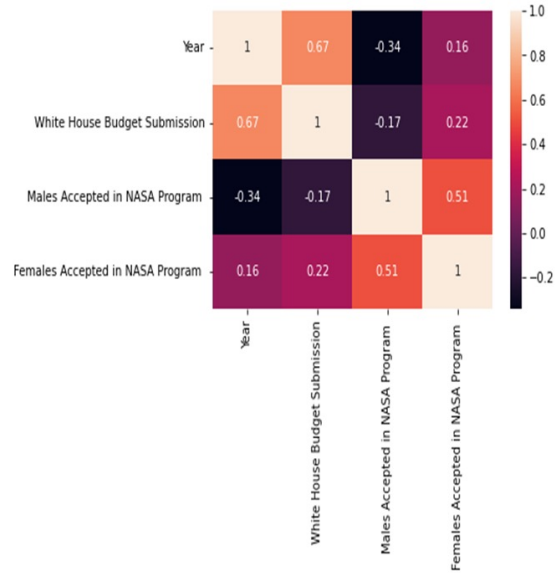
```
Ttest_1sampResult(statistic=-4.706948275639553, pvalue=1.2796410491460158e-05)
```

1. The null hypothesis , mean = 50
2. Alternaye hypothesis , the mean is not equal to 50 results. The p-value is less than 0.05 and the null hypothesis is rejected and the alternate hypothesis is correct.

Correlation Plot

```
[ ] #Correlation Plot
    import seaborn as sns
    sns.heatmap(dataframe.corr(), annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f1f8f2d8690>
```



The correlation between the budgjet and females, and budget and males is a weak corelation. The budget has nothing todo with how many females and males are accepted in the NASA program.

Linear Regression of Females vs. Budget

```
[ ]  #Linear Regression of Females versus Budget
     %matplotlib inline
     import  statsmodels.api as sm
     import statsmodels.stats.api as sms
     from scipy.stats import boxcox
     x = dataframe['Females Accepted in NASA Program ']
     y = dataframe['White House Budget Submission']
     model = sm.OLS(x,y).fit()
     model.summary()
```

OLS Regression Results

Dep. Variable:	Females Accepted in NASA Program	**R-squared (uncentered):**	0.207
Model:	OLS	**Adj. R-squared (uncentered):**	0.195
Method:	Least Squares	**F-statistic:**	17.75
Date:	Tue, 17 May 2022	**Prob (F-statistic):**	7.58e-05
Time:	00:49:26	**Log-Likelihood:**	-259.61
No. Observations:	69	**AIC:**	521.2
Df Residuals:	68	**BIC:**	523.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
White House Budget Submission	0.0004	9.85e-05	4.213	0.000	0.000	0.001

Omnibus:	40.985	**Durbin-Watson:**	2.321
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	100.400
Skew:	2.003	**Prob(JB):**	1.58e-22
Kurtosis:	7.344	**Cond. No.**	1.00

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

R-square values shows that the budget has nothing to do with the budget assigned. The value of R-square is 0.2, which means it is a weak correlation .

Linear Regression for Year and Budget

```
#Linear Regression for Year and Budget
%matplotlib inline
import  statsmodels.api as sm
import statsmodels.stats.api as sms
from scipy.stats import boxcox
x = dataframe['Year']
y = dataframe['White House Budget Submission']
model = sm.OLS(x,y).fit()
model.summary()
```

OLS Regression Results

Dep. Variable:	Year	**R-squared (uncentered):**	0.695
Model:	OLS	**Adj. R-squared (uncentered):**	0.691
Method:	Least Squares	**F-statistic:**	155.1
Date:	Tue, 17 May 2022	**Prob (F-statistic):**	3.27e-19
Time:	00:56:49	**Log-Likelihood:**	-581.14
No. Observations:	69	**AIC:**	1164.
Df Residuals:	68	**BIC:**	1167.
Df Model:	1		
Covariance Type:	nonrobust		

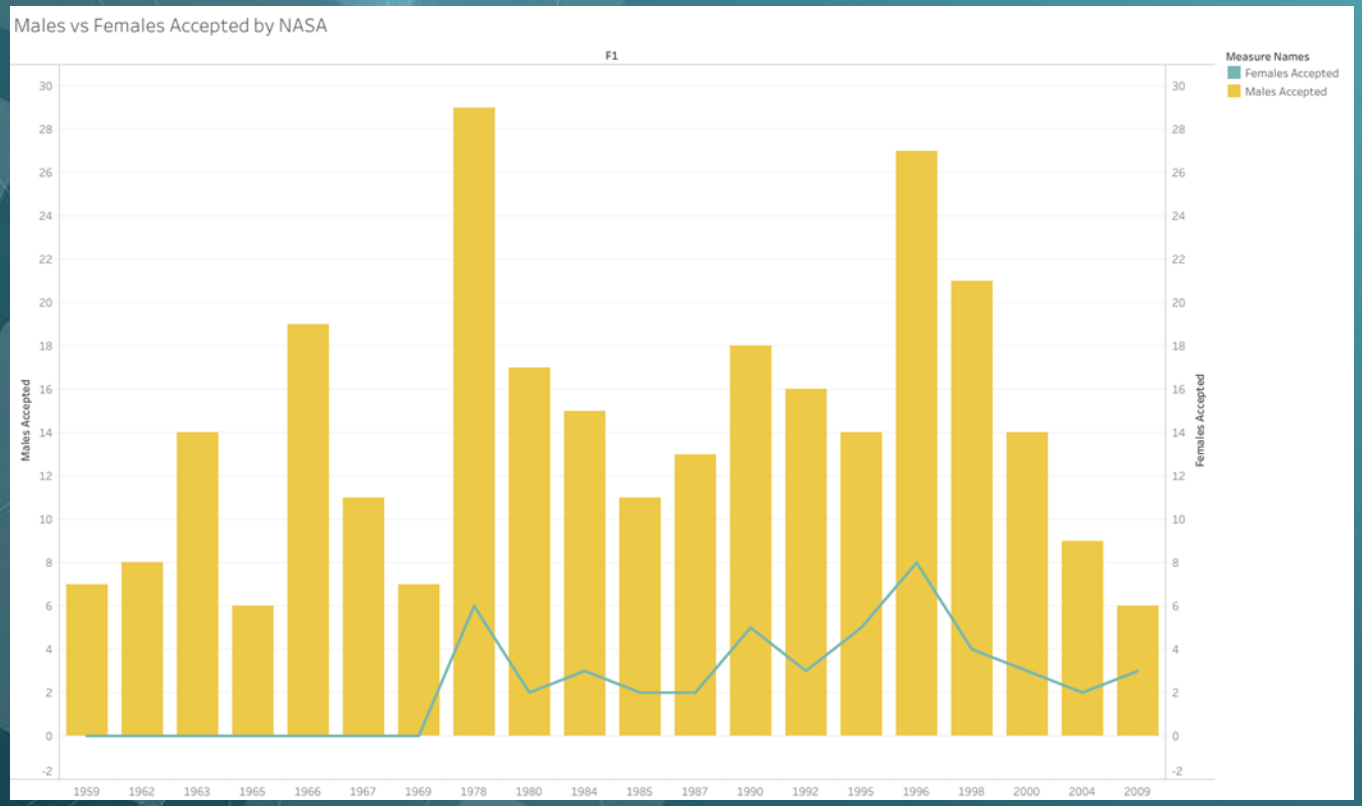
	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
White House Budget Submission	0.1295	0.010	12.455	0.000	0.109	0.150

Omnibus:	13.866	**Durbin-Watson:**	0.146
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	3.886
Skew:	-0.164	**Prob(JB):**	0.143
Kurtosis:	1.885	**Cond. No.**	1.00

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

The Year does have an affect on the Budget by looking at the R-square value which is 0.7. The higher the R-square value the higher the correlation.

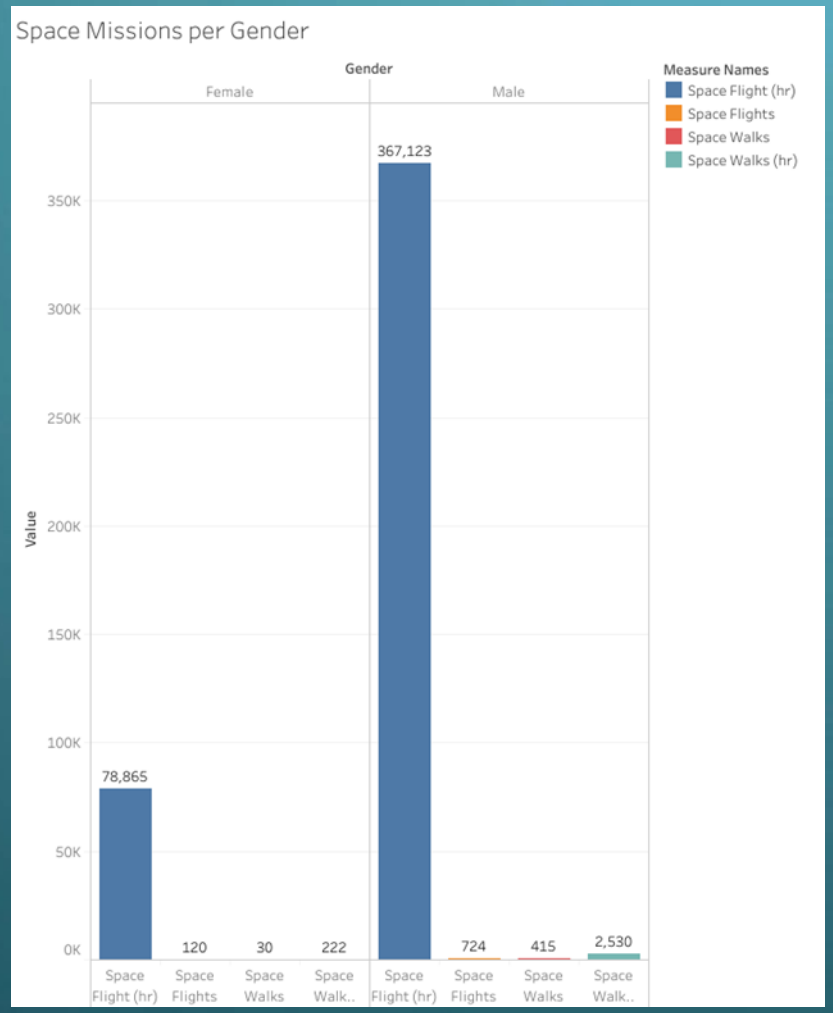
Results: Hiring Rates between Males and Females in NASA



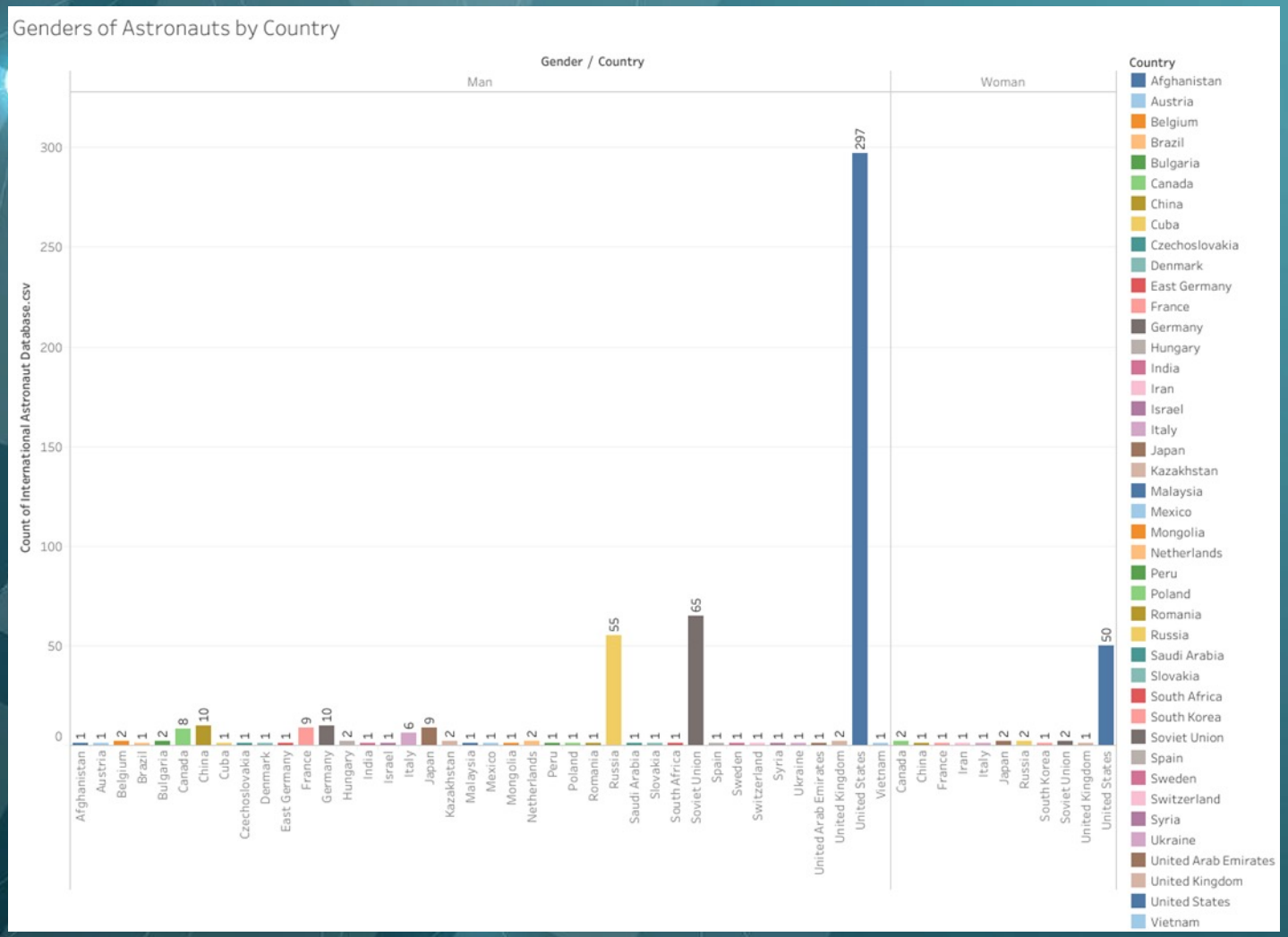
- When the NASA program was first started in 1959, it was only populated by males who had military background.
- In 1978, the first females were selected as astronaut candidates
- Every year since 1978, the intake of females into the program are always less than those of males accepted into the NASA program

Results: Ratio of males to females that are sent on space missions

- Males had a greater number of collective hours on Space flights and Space Walks.
- Females had a higher average of individual hours!

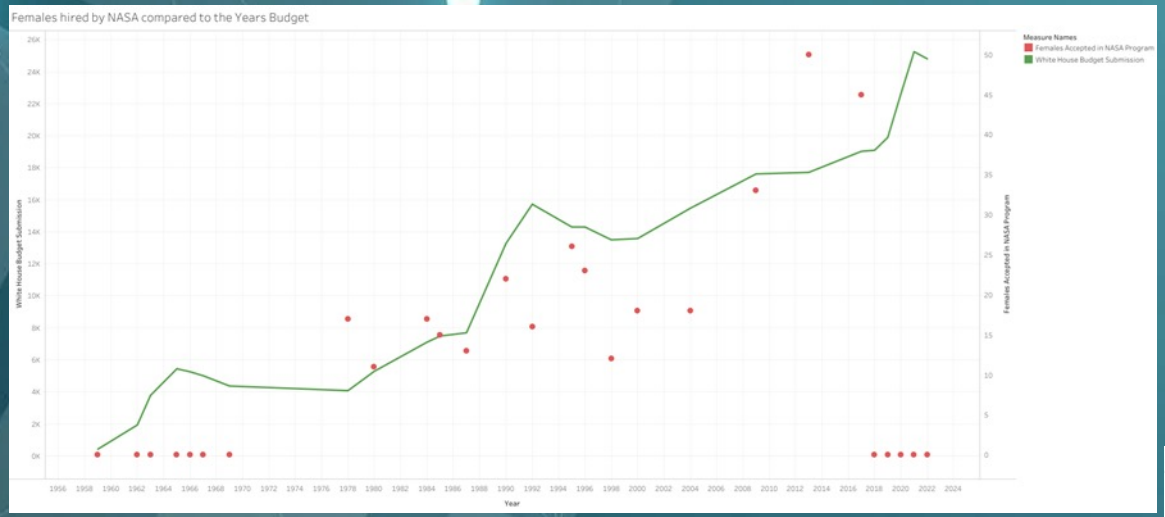


Results: Exploratory Findings



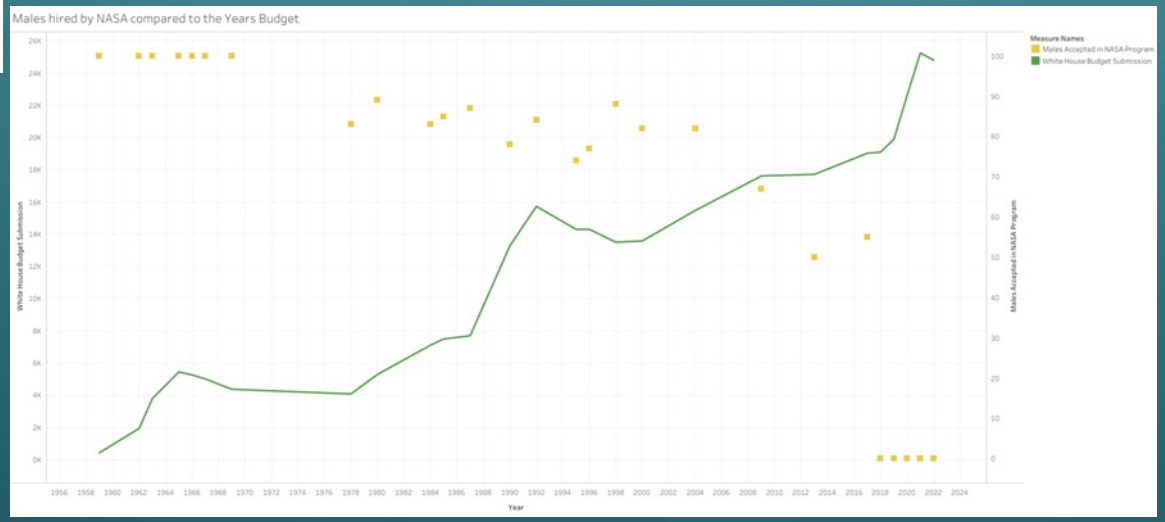
- The United States produces the most astronauts
- The Soviet Union and Russia produce the second and third highest number of astronauts
- Only 9 countries hired both male and female astronauts
- 2 countries (Iran and South Korea) only have females representing their countries as astronauts

Results: Exploratory Findings



- Budget has no influence on the hiring of female astronauts

- Budget has no significant influence on the hiring of male astronauts



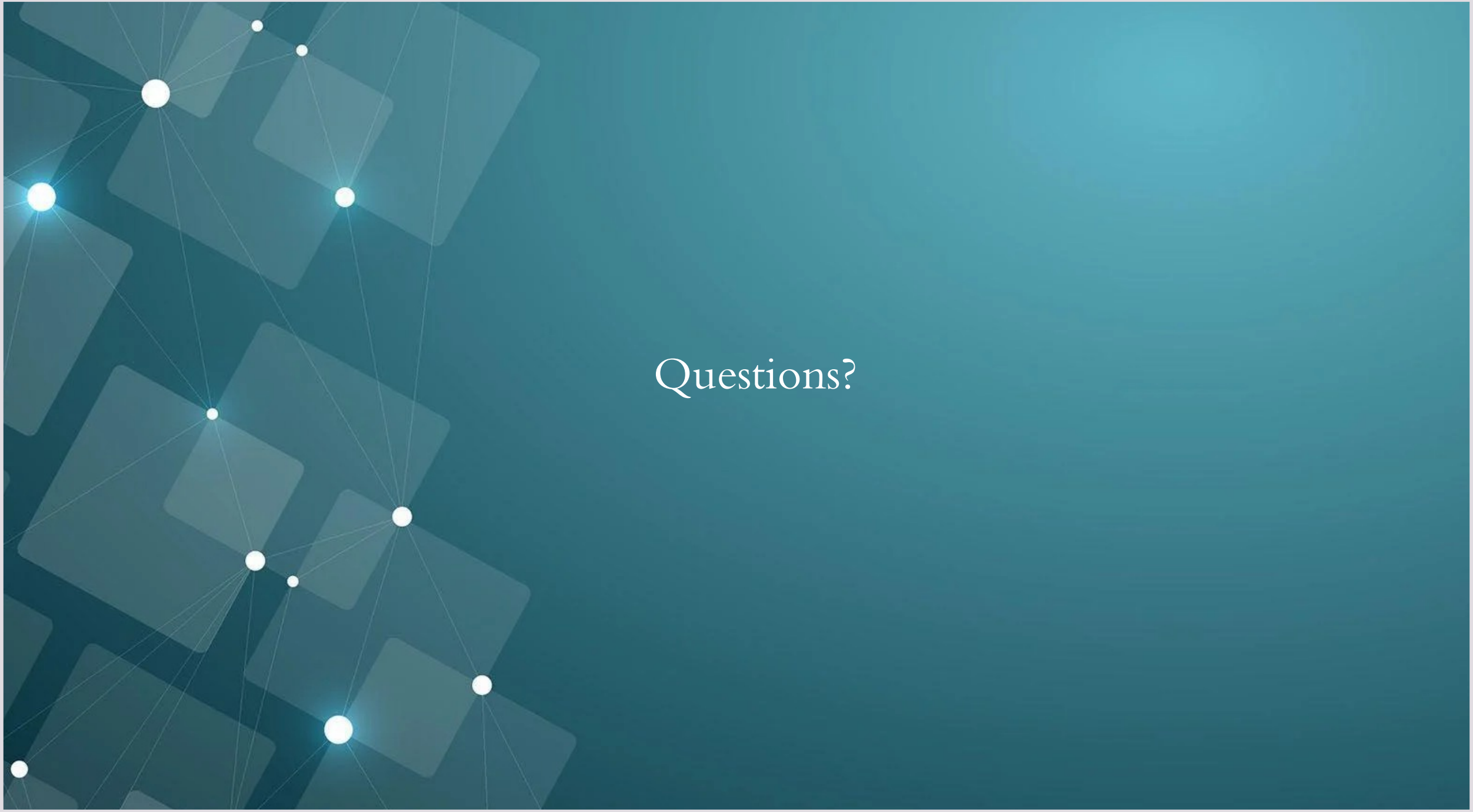
Summary

- The US is the leading country in having astronauts in both men and women
- In the US, NASA has only 10% women astronauts
  In the world women astronauts consist of only 14%
- One question that we asked ourselves is “Did the white house budget for the NASA program have an impact on how many women and men were hired into it?”

Conclusion

How did your findings impact the world at large?

What's important about this work?


Questions?