Data Science & Visualization Project 1

HOSPITAL CHARGE DATA

NOREEN MALIK, RENE VENEMA ("DUTCH"), JOHN WARLOP & ALEX WOOTEN

FABULOUS FOUR

Motivations and Summary

- **HYPOTHESIS**
  - **Higher adjusted gross income, higher out of pocket costs**
  - **Population does not effect cost**
  - **Correlation between Average Covered Charges and Average Medicare Payments**

Data

Sources

- **DATA.CMS.GOV (main date set):**

  https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-_Payment-System-IPPS-Provider/97k6-zzx3/data

- **IRS:**

  https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2015-zip-code-data-soi

- **Population data:**

  www.census.gov

Clean Up/Exploration

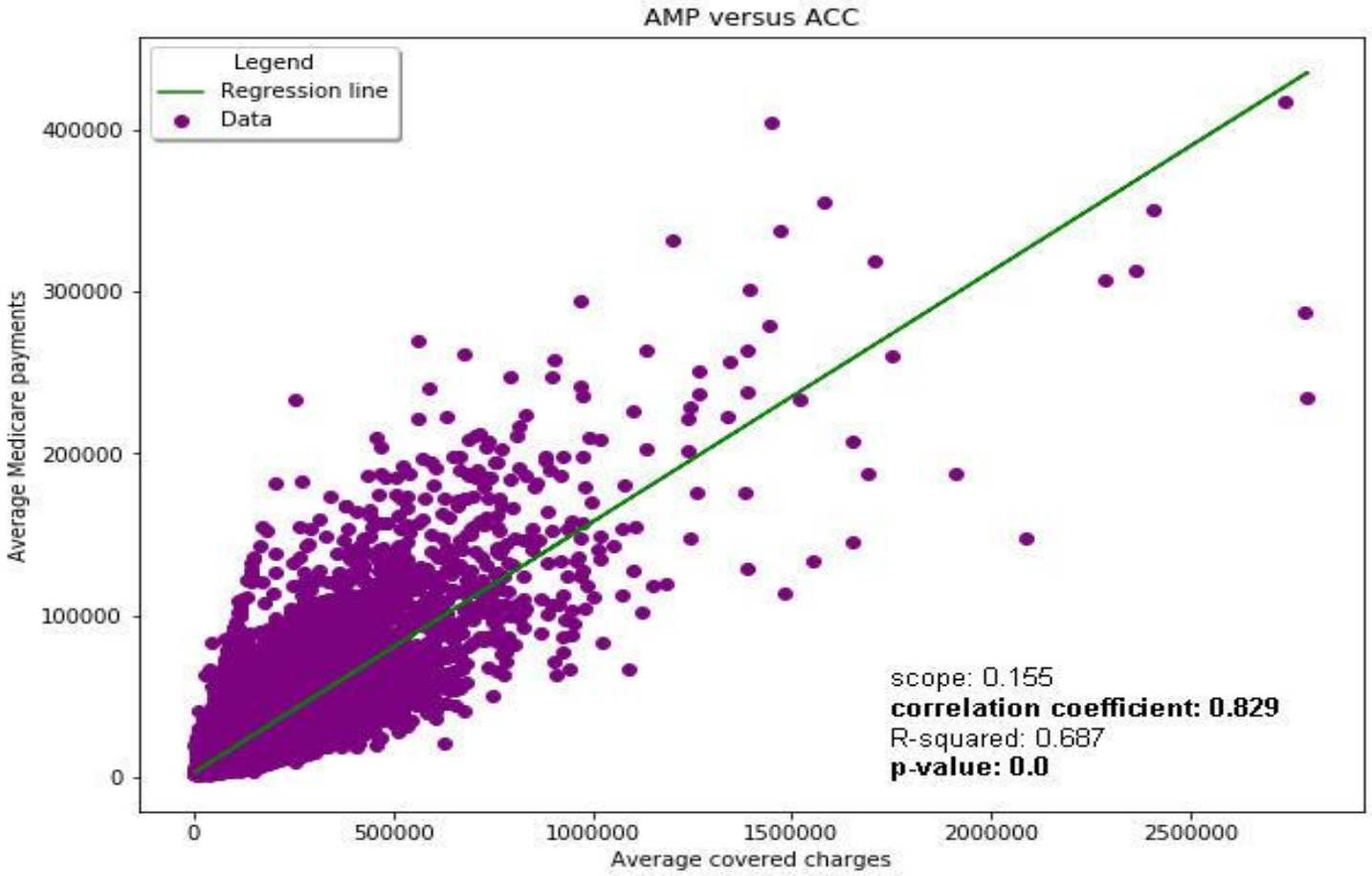
Data Cleansing

- **data.cms.gov (main data set)**

  Data imported using Sodapy was all strings. Changed data type for columns involved to float and integer. No null values/missing data in data set.

- **IRS data- Needed to discard irrelevant data and extract only the needed info to data frame, renamed the columns for better understanding, and ultimately created a pandas CSV file with the data extracted.**
- **Population data-There were different formats of state columns that needed to be merged together. Found a state abbreviation table online, which allowed the data frames to be merged**

Average Medicare Payments versus Average Covered Charges



Hypothesis

Hypothesis

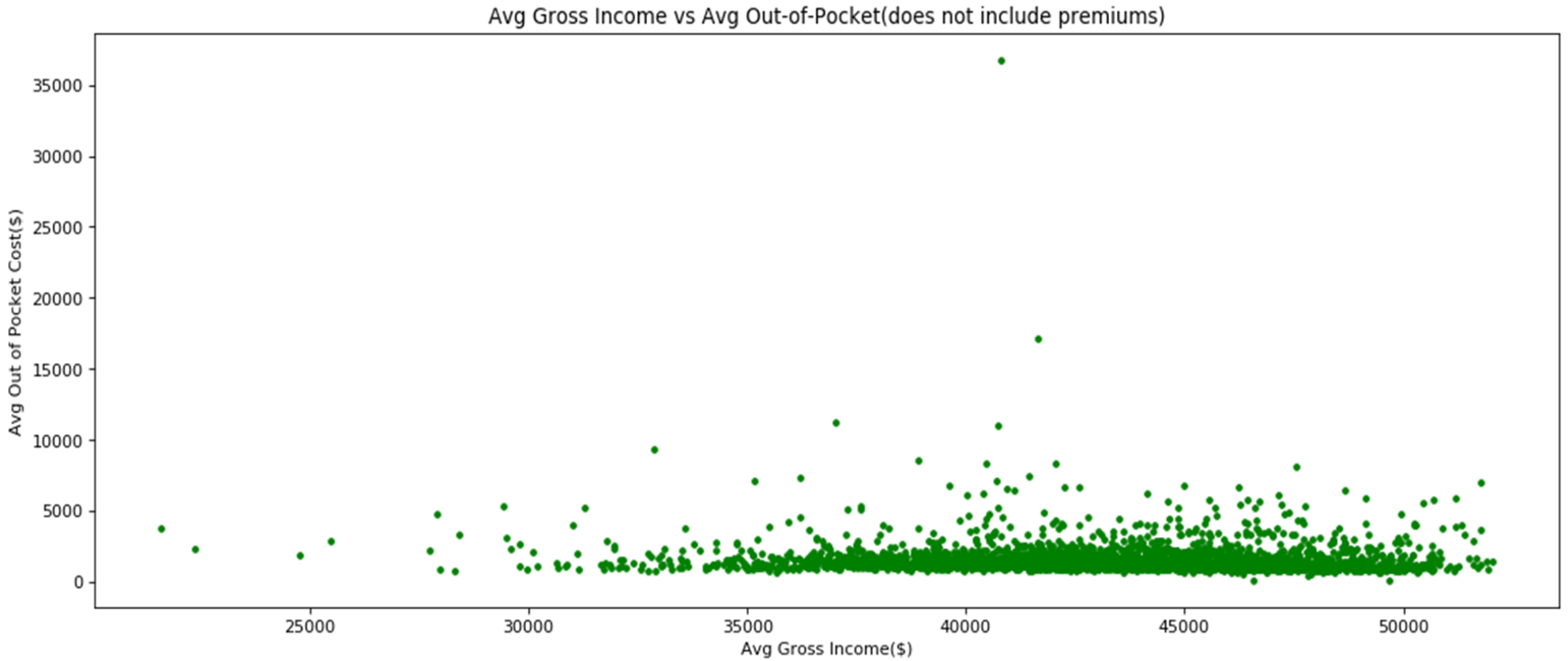
Medicare recipients with higher adjusted gross incomes pay higher out of pocket costs

Assumptions

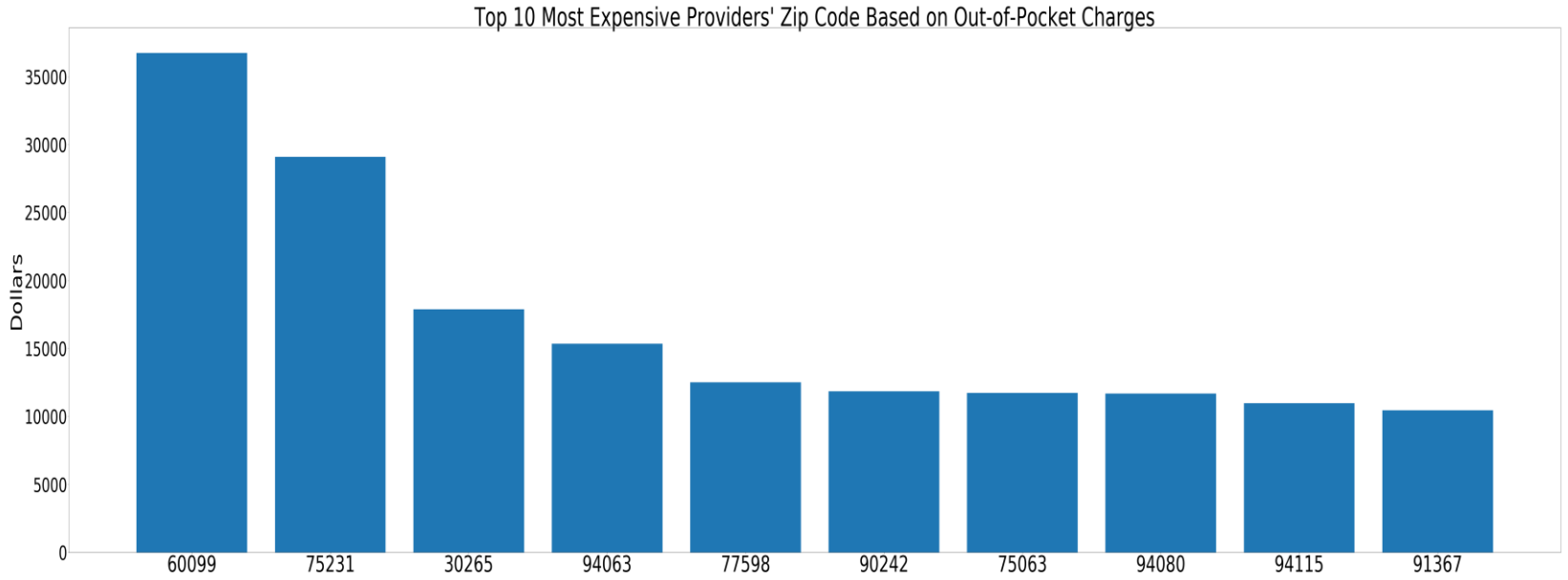
Assumptions

- IRS data was in 25K bins. I assumed all filers had AGI midpoint of 25K bins
- Joint filers counted as two people and single filers filers as one person
- Out-of-pocket cost is total_payments - total_medicare payments(actual oop costs is difficult to calculate)

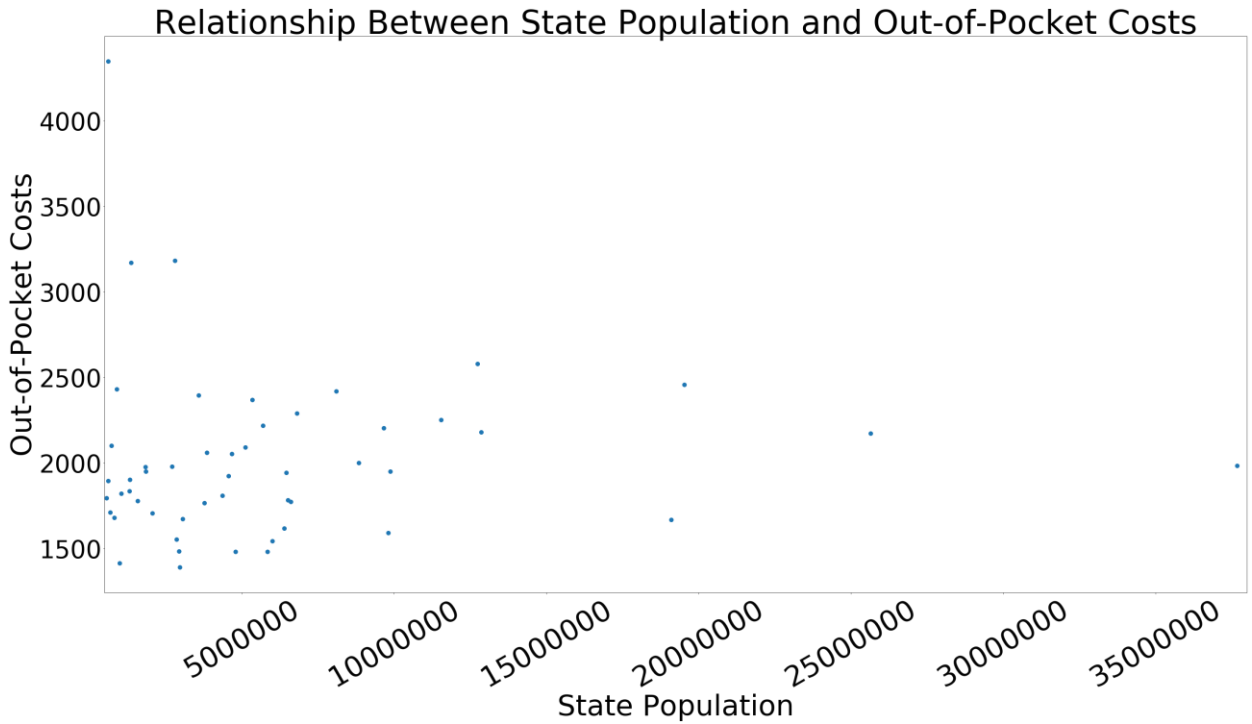
Average Gross Income VS Out of Pocket Costs



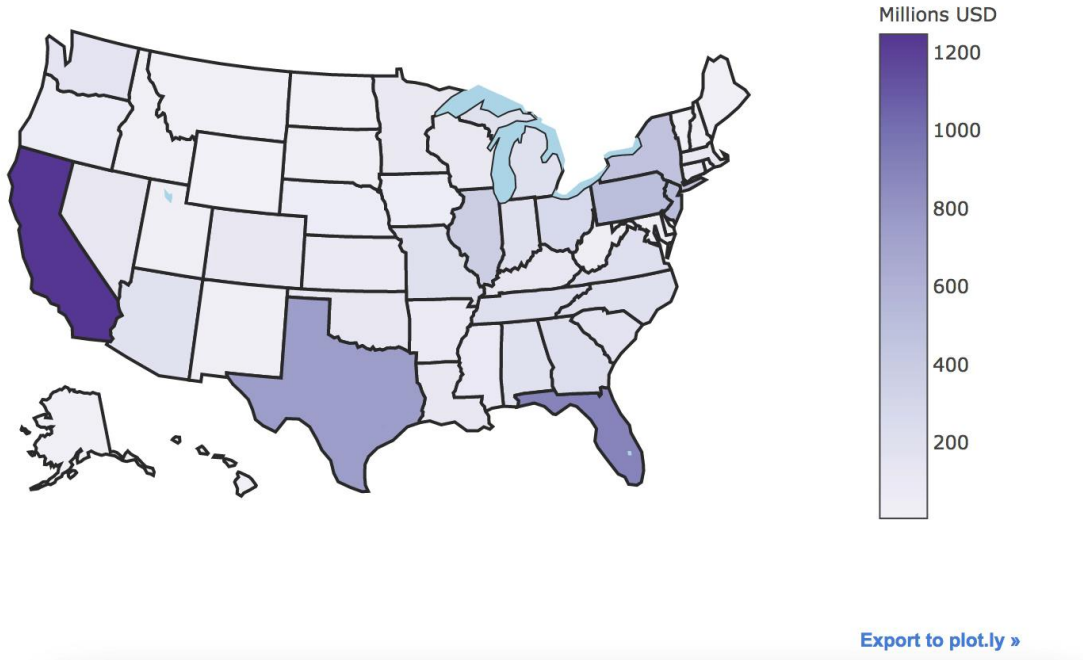
Most Expensive Zip Codes



State Population VS Out of Pocket Costs



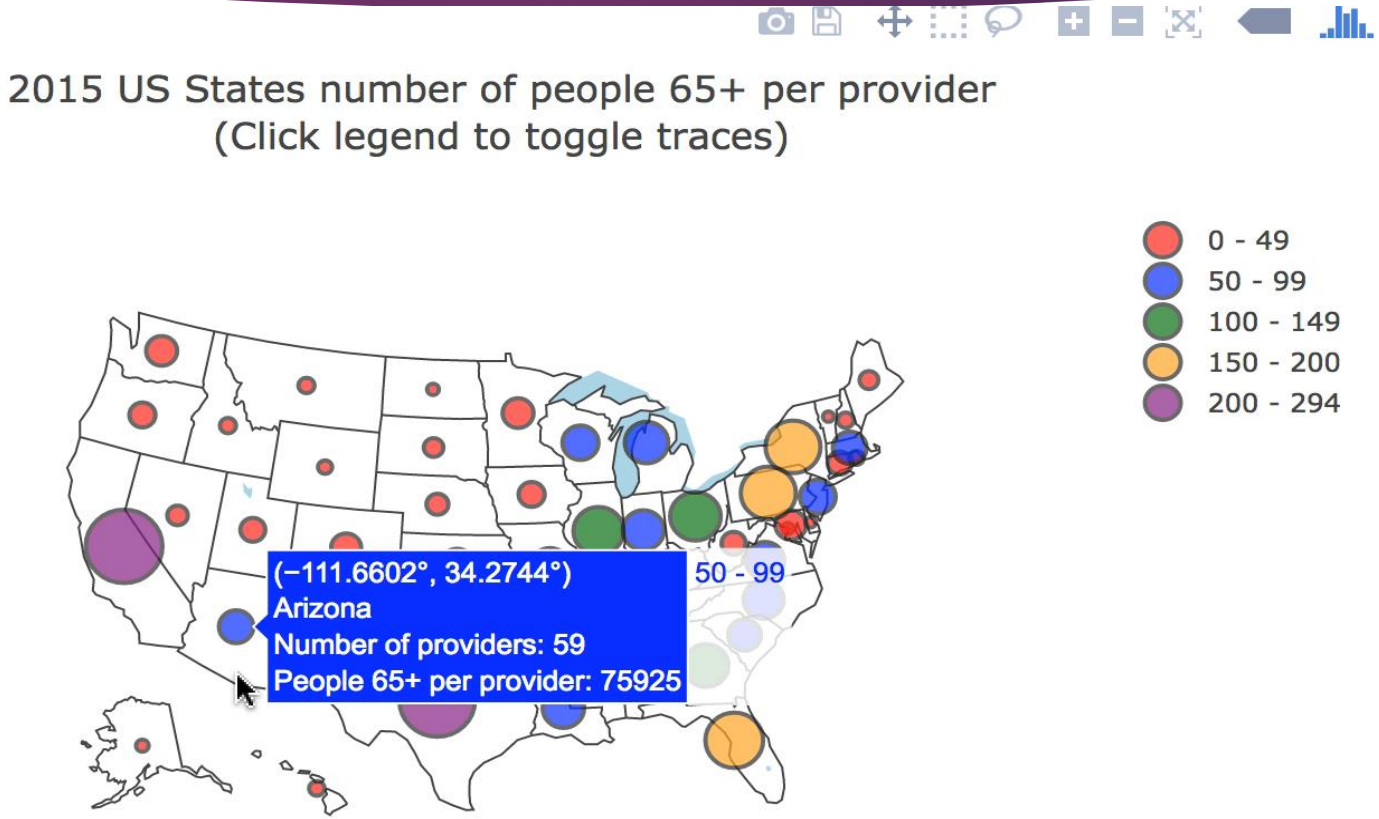
2015 US States; overall provider deficits



Top 5 2015 US States with largest deficit

	State	Total providers	ACC Total	AMP Total	ATP Total	Payment difference	Deficit
50	WY	10	9.44	3.05	3.53	0.48	6.39
46	VT	6	10.86	4.38	5.04	0.65	6.47
20	MD	47	84.79	71.62	78.59	6.97	13.17
11	HI	12	20.76	6.62	8.03	1.41	14.14
26	MT	13	21.21	6.42	7.29	0.87	14.79

2015 US States; population (65+) per no. of providers



Top 5 2015 US States with highest number of people 65+ per provider

	name	Population	State	total_providers	People_65+_per_provider
8	Delaware	641516	DE	6	106919
9	Florida	15764516	FL	167	94398
47	Washington	4145168	WA	48	86358
31	New Jersey	5374708	NJ	64	83980
33	Nevada	1687040	NV	21	80335

Assumptions

- **Assumptions**
- **Out of Pocket = Total cost-Medicare payment**
- **Assumed joint file taxes were 2 people, where single was one person.**
- **Had to assume adjusted gross income based on bins**

Conclusions

- Discovered overall AGI does not effect cost of procedure. Premiums effect how much you pay out of pocket, but AGI does not effect cost of procedure.
- Populations does effect cost in that higher population, higher cost
- Pretty strong correlation between cost and Medicare payments
- Difficulties:

  Understanding GIT. All new to GIT spent more time on this than anticipated.

  Understanding definitions (Average Coverage Charges, Average total payments,

  Average Medicare payments).
- If we had more time we would focus on trying to dive deeper with our income analysis, needed more time to look for data that would be specific enough. AGI per zip code. Not want to make too many assumptions
- More interactive plots.
- We did not have a hypothesis before analyzing data, the data directed our project.

Q&A

THANK YOU!!