US Demographics and Public Resources

—

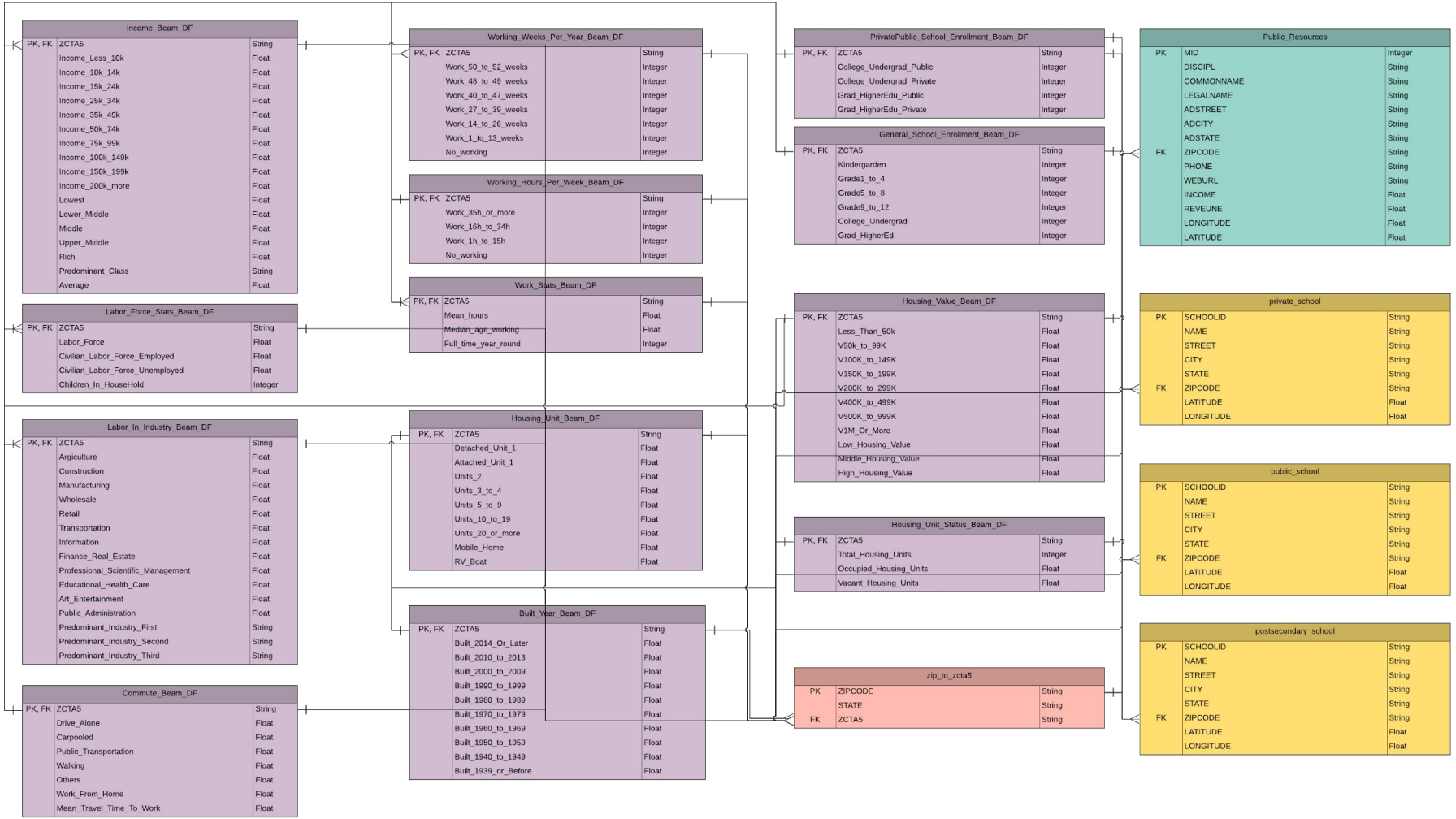
Sashimi

Chosen Datasets - Primary

- American Housing Survey
  - Basic demographic, household income
  - Housing status and information
  - Predominant industry, employment rate, average working hour/week
  - Overall education level, school enrollment rate
  - Location in ZCTA5

Chosen Datasets - Secondary

- Institute of Library and Museums
  - Discipline, income, revenue
  - Location in ZIPCODE, longitude, latitude
- Education Demographic and Geographic Estimates
  - Public, private and postsecondary school
  - Location in ZIPCODE, longitude, latitude
- UDS Mapper
  - Location in ZIPCODE
  - Location in ZCTA5


Income_Beam_DF
PK, FK ZCTA5 String
Income_Less_10k Float
Income_10k_14k Float
Income_15k_24k Float
Income_25k_34k Float
Income_35k_49k Float
Income_50k_74k Float
Income_75k_99k Float
Income_100k_149k Float
Income_150k_199k Float
Income_200k_more Float
Lowest Float
Lower_Middle Float
Middle Float
Upper_Middle Float
Rich Float
Predominant_Class String
Average Float
Labor_Force_Stats_Beam_DF
PK, FK ZCTA5 String
Labor_Force Float
Civilian_Labor_Force_Employed Float
Civilian_Labor_Force_Unemployed Float
Children_In_HouseHold Integer
Labor_In_Industry_Beam_DF
PK, FK ZCTA5 String
Argiculture Float
Construction Float
Manufacturing Float
Wholesale Float
Retail Float
Transportation Float
Information Float
Finance_Real_Estate Float
Professional_Scientific_Management Float
Educational_Health_Care Float
Art_Entertainment Float
Public_Administration Float
Predominant_Industry_First String
Predominant_Industry_Second String
Predominant_Industry_Third String
Commute_Beam_DF
PK, FK ZCTA5 String
Drive_Alone Float
Carpooled Float
Public_Transportation Float
Walking Float
Others Float
Work_From_Home Float
Mean_Travel_Time_To_Work Float
Working_Weeks_Per_Year_Beam_DF
PK, FK ZCTA5 String
Work_50_to_52_weeks Integer
Work_48_to_49_weeks Integer
Work_40_to_47_weeks Integer
Work_27_to_39_weeks Integer
Work_14_to_26_weeks Integer
Work_1_to_13_weeks Integer
No_working Integer
Working_Hours_Per_Week_Beam_DF
PK, FK ZCTA5 String
Work_35h_or_more Integer
Work_16h_to_34h Integer
Work_1h_to_15h Integer
No_working Integer
Work_Stats_Beam_DF
PK, FK ZCTA5 String
Mean_hours Float
Median_age_working Float
Full_time_year_round Integer
Housing_Unit_Beam_DF
PK, FK ZCTA5 String
Detached_Unit_1 Float
Attached_Unit_1 Float
Units_2 Float
Units_3_to_4 Float
Units_5_to_9 Float
Units_10_to_19 Float
Units_20_or_more Float
Mobile_Home Float
RV_Boat Float
Built_Year_Beam_DF
PK, FK ZCTA5 String
Built_2014_Or_Later Float
Built_2010_to_2013 Float
Built_2000_to_2009 Float
Built_1990_to_1999 Float
Built_1980_to_1989 Float
Built_1970_to_1979 Float
Built_1960_to_1969 Float
Built_1950_to_1959 Float
Built_1940_to_1949 Float
Built_1939_or_Before Float
PrivatePublic_School_Enrollment_Beam_DF
PK, FK ZCTA5 String
College_Undergrad_Public Integer
College_Undergrad_Private Integer
Grad_HigherEdu_Public Integer
Grad_HigherEdu_Private Integer
General_School_Enrollment_Beam_DF
PK, FK ZCTA5 String
Kindergarden Integer
Grade1_to_4 Integer
Grade5_to_8 Integer
Grade9_to_12 Integer
College_Undergrad Integer
Grad_HigherEd Integer
Housing_Value_Beam_DF
PK, FK ZCTA5 String
Less_Than_50k Float
V50k_to_99K Float
V100K_to_149K Float
V150K_to_199K Float
V200K_to_299K Float
V400K_to_499K Float
V500K_to_999K Float
V1M_Or_More Float
Low_Housing_Value Float
Middle_Housing_Value Float
High_Housing_Value Float
Housing_Unit_Status_Beam_DF
PK, FK ZCTA5 String
Total_Housing_Units Integer
Occupied_Housing_Units Float
Vacant_Housing_Units Float
zip_to_zcta5
PK ZIPCODE String
STATE String
FK ZCTA5 String
Public_Resources
PK MID Integer
DISCIPL String
COMMONNAME String
LEGALNAME String
ADSTREET String
ADCITY String
ADSTATE String
FK ZIPCODE String
PHONE String
WEBURL String
INCOME Float
REVEUNE Float
LONGITUDE Float
LATITUDE Float
private_school
PK SCHOOLID String
NAME String
STREET String
CITY String
STATE String
FK ZIPCODE String
LATITUDE Float
LONGITUDE Float
public_school
PK SCHOOLID String
NAME String
STREET String
CITY String
STATE String
FK ZIPCODE String
LATITUDE Float
LONGITUDE Float
postsecondary_school
PK SCHOOLID String
NAME String
STREET String
CITY String
STATE String
FK ZIPCODE String
LATITUDE Float
LONGITUDE Float

Area of Interest

- Average housing value **vs** average household income
- Industry vs average household income

- Which area has better public resources **vs** which area is lacking
- Investment vs demand on education resources
- Where public resources located at **vs** living condition nearby

Problems and Solutions

- Massive Table
- Column names
- Different location keys (Zip Code vs. ZCTA5)

- Omit unnecessary columns and split up table for normalization
- Beam transform
- UDS Mapper to connect

Pipelines

- SQL
  - Cast special null markers as ‘Null’
  - Update schema type
- Beam
  - Update column names from code to human readable name
  - Generate columns through computation
- DAG

Live Demo

- Income: generating average and assigning socioeconomic classes
- Cross joins
- Visualization

Future Improvements

- More comprehensive “public resources”
  - Combine Schools with Libraries and Museums
- Collect some data on government funding
- Zip Code might be too granular when aggregating data
  - Example: I live in Riverside but go to UT
- Better estimate on the needed data to avoid unnecessary work

Thanks & Questions