**Questions for Data Skills Sharing Meeting with Transportation**

**General Questions**

We have general questions about the when and why of using different open source programming languages to enhance data collection, exploration, modeling, and visualization. We are, however, open to other suggestions that are currently not on our radar, but you find valuable for working with and presenting data.

* Why and when would you choose Python vs R?
* Do you have tips for integrating Python with Excel (or other Microsoft Office products) to automate work and enhance analysis?
* Is it a smooth process to export Python and R outputs to a Microsoft Office environment such as what was done for your weekly reports? If not, what challenges should we expect to encounter?
* How involved is it to use screen (or data) scraping for extracting data from websites such as what was done with DMV data?

**Projects**

* **Federal Microdata**

Working with federal microdata such as ACS PUMS and CPS data are foundational to all our projects. We are looking for ways to enhance our data collection and analysis capabilities.

* + Which programming language do you use to work with microdata?
  + From which site do you import microdata (e.g., IPUMS, Census Bureau, etc.) and how?
  + Do you calculate estimates beyond count statistics (e.g., medians, averages, etc.)? If so, which program language do you prefer to do it in and why?
  + Which programming language do you use to calculate standard error using replicate weights and why? Are these calculated for count as well as descriptive statistics?
  + Do you conduct cloud computing for memory intensive processes? If so, could you give a brief overview of what is involved?
* **Interactive Reports**

Our division has been exploring different web mapping and data visualization platforms to bring a fresh approach to presenting our data. Our [Migration Brief](https://www1.nyc.gov/assets/planning/download/pdf/planning-level/housing-economy/migration-info-brief.pdf?r=1) is an example of a typical static document we would like to enhance by transforming it into an [interactive report](https://migration.planning.nyc.gov/#age) that we can share. We have been testing different ways for [interactive report](https://nyco365.sharepoint.com/:u:/r/sites/NYCPLANNING/population/Shared%20Documents/08%20Data%20Viz/00%20R/04%20Working/BriefLayout.html?csf=1&web=1&e=sa8WAa) using R (the html file is on SharePoint and requires you download it to your desktop and then open in your browser).

* + What options are available for creating reports with interactive charts, graphs, and maps? We are looking for ways to allow users to explore and filter data on graphs and tables, select specific categories on a legend, etc.
  + Do you have a preferred site for hosting data (e.g., GitHub, Shiny, etc.)? What are the advantages of one site over the other?

**Knowledge Transfer**

* Could you create a library on SharePoint where we can share R and Python scripts and reproducible examples to help us get started?
* How would shadowing work?