CTPP Status Report

U.S. Department of Transportation
Federal Highway Administration
AASHTO Committee on Planning

AASHTO
THE VOICE OF TRANSPORTATION

TRB Census Subcommittee
Bureau of Transportation Statistics
Federal Transit Administration

Census Transportation Planning Products (CTPP) Updates

Penelope Weinberger, AASHTO, pweinberger@aashto.org

Hello for the first time in two years! A lot has happened since we last published a newsletter. The 2022 Census Data Transportation Conference in Reno was a highlight, while getting MacroSys onboard to develop the next CTPP data access software was another. The new data access software will include an application programming interface (API) that will enable users to further customize CTPP outputs.

The CTPP Oversight Board is moving forward with new states and MPOs coming on. The Board is about to seek our next round of funding! (Keep an eye open for that, and if you work for a state DOT, please let them know!)

Our online training was a success, though it is on hiatus until the new software comes. The new software is anticipated in mid to late 2023. We are close to an agreement with the Census Bureau on our next CTPP dataset. Please see Clara's article for a lot more detail on that.

We hope to see you at the TRB Innovations in Travel Analysis and Planning Conference *(ITAP)* in June in Indianapolis, IN. Things are always changing so we encourage you to drop by our table and see what's new.

Stay safe, be well, and use good data.

Editor's note: In case you were not at ITAP, rebranding of all AASHTO products has resulted in a new name. CTPP will now be known as AASHTO Census Transportation Solutions, or ACTS!

Contents

2017-2021 CTPP Table Request from the Census Bureau

Clara Reschovsky, USDOT Bureau of Transportation Statistics, clara.reschovsky@dot.gov

Over the last 18 months, the AASHTO CTPP Oversight Board's Table Subcommittee has been working with the Census Bureau on the next CTPP data package based on 2017-2021 American Community Survey (ACS) data. As of January 2023, we have "approval with restrictions," which is a major step forward.

The initial concerns from Census Bureau's Disclosure Review Board (DRB) on the original package request, which included everything we could think of asking for since there were no preliminary parameters of what would be acceptable, were:

1. Reduce the overall size of the package.
2. Standardization of Universes, Variables, and Categories wherever possible to match standard Census ACS products.
3. Perturbation techniques must be employed to account for unexpected deviations in results *(see below for further discussion of perturbation techniques)*.
4. Any flow must be subject to an unweighted minimum of 3 observations. This means 3 observations had to respond to the ACS that they live in one geography and work in the other.
5. Non-residence geographies must have 50 unweighted cases to avoid suppression. This means that part 2 *(workplace)* data at small geographies will be in jeopardy of being suppressed.
6. Mean, Median, and Aggregate tables must have a minimum of 3 observations to be tabulated.

Response Strategy

With this in mind, the Table Subcommittee developed a response to the Census Bureau (CB) to determine a path forward. The Table Subcommittee worked with industry and subject matter experts on the response. The strategy included:

1. Eliminating block groups in part 3 *(Flows)* and many of parts 1 *(Residence)* and 2 (*Workplace*). *Retaining a very small list of univariates in parts 1 and 2 tables at the block group level.*
2. Standardizing universes, variables, and categories/cohorts to the extent possible.
3. Developing a sound justification for keeping CTPP-specific universes, variables, or categories.
4. Expecting perturbation decisions to be data driven and originate at the CB.
5. Documenting the loss of data that occurs when the rule of 3 observations for flows is implemented.
6. Mean and Aggregate tables with less than 3 observations are not very meaningful and are a disclosure risk.

Restrictions

The Census Bureau has reviewed the proposed solutions and the DRB issued its decision to approve the package, but to require some changes to the processing. These rules replace the original response from above. These current restrictions include:

1. When using the standard ACS microdata if a table is not available for a requested geography, it will be run using the perturbed microdata.
2. If a table category is custom to CTPP and is not available or not a combination of a standard ACS published table category, the table will be subject to the Rule of 5 observations using standard ACS Microdata or Rule of 3 observations using perturbed microdata.

Back to Top

3. For tables with 100 cells or more, small geographic entities will not be produced unless a suitable collapse can be created with fewer than 100 cells.
    a. For tables with *100-125 cells*, only places with *10,000 or more population and higher-level* geographies will be produced.
    b. For tables with *more than 125 cells*, only geographies with *100,000 or more population* will be produced.
4. Mean, Median, and Aggregate tables must have a minimum of 3 observations to be tabulated.
5. ACS unweighted counts *(only produced for the Nation, State, County and Place for the 1-Year and 5-Year)* were protected starting in the ACS 2019 and 2015-2019 5-Year, and moving forward, using differential privacy by adding noise generated from the Laplace mechanism *(i.e., a symmetric version of an exponential distribution)*.
6. All cells in any American Community Survey special tabulation must be rounded and the rounding schema is the same as in previous CTPP packages.
7. The CTPP variable "Age of youngest person in the household" is no longer allowed, due to data disclosure concerns.

Perturbation is required to produce the CTPP tables similarly as in the previous CTPP tabulations. Perturbation is a technique that is performed on microdata records prior to producing a frequency of cross-tabulating variables. Specifically, perturbation swaps values in one or more locations for variables on the microdata records. This introduces noise and uncertainty into the resulting tabulation that masks the identity of ACS respondents. To limit the "noise" there are bounds on the geographic areas in which the swapping occurs. Additionally, the weights are readjusted to preserve the integrity of the resulting weighted counts.

Implications for CTPP

In looking very closely at the implications of these restrictions, it is anticipated that the 2017-2021 CTPP table set will be different from previous CTPP table sets. Many of the changes are necessary due to increased concerns about disclosure of ACS respondents' identities.

The first two restrictions (above) are the most impactful on CTPP. Since perturbation will be pervasive throughout all three parts of the CTPP, the AASHTO Oversight Board of CTPP has decided to ask the Census Bureau to perturbate

> In addition to the ongoing concerns about small geographies, there are also concerns about demographic slivers. Demographic slivers are of particular concern in CTPP when the custom categories split into a standard ACS category, such as the Age category of 15–17-year-olds that gets split apart by CTPP workers who are defined as 16 years and older.

the entire package. This will make the data tables more consistent with one another. Once perturbation has been performed, many tables will still be subject to suppression as well. The suppression will affect perturbated tables where there are three or fewer unweighted observations for a particular geographic entity in ACS. We are hopeful that perturbation can help with limiting some of the suppression.

Restriction 3 is causing the team to go back to the large tables and reevaluate the tables to see if additional collapses would work better to allow the production of those tables for smaller geographies. The tables would still be subject to the perturbation and suppression rules discussed above. The team is exploring asking Census to tabulate multiple versions, the first would be the large cell count version which would only be available for large geographies, but then a collapsed version would be tabulated and available for all geographic levels in CTPP. The team is currently working on figuring out how to collapse the variables in a logical manner to preserve the data utility.

For restrictions 4-6, the impact on the CTPP is as expected. Any mean, median, or aggregate is not particularly meaningful with few observations, and rounding has been in place with CTPP since 2000. As for the unweighted counts, the Census Bureau has moved away from producing unweighted counts due to concerns about protecting the identities of respondents. Finally, restriction 7 has been a useful variable in predicting behavior among households with children. The Census Bureau tabulates the Presence and Age of Children in Family Households as seen in ACS standard table B11003. We are working with Census to ascertain if this variable will be sufficiently suitable for transportation planning uses.

In conclusion, the Table Subcommittee of the AASHTO Oversight Board for CTPP is continuing to work with the Census Bureau's DRB and special tabulation coordinator to secure the most useful package possible for the 2017-2021 CTPP.

If you have questions, concerns, or comments about the package, please reach out to:

- Penelope Weinberger - *pweinberger@aashto.org*
- Arash Mirzaei - *amirzaei@nctcog.org*
- Clara Reschovsky - *clara.reschovsky@dot.gov*

New CTPP Data Access Software is On the Way
Colby Brown, Manhan Group, colby@manhangroup.com

In early 2022, the American Association of State Highway and Transportation Officials (AASHTO) began working with MacroSys LLC and Manhan Group to develop a new system of web-based tools for accessing data from the Census Transportation Planning Products (CTPP) program. Specifically, the new system will be comprised of three enhanced components: a front-end data portal (the graphical user interface, or GUI), a back-end database, and a web application programming interface, or API. The stated goal of the project is to deliver an updated data access solution reflecting current technology and user needs.

User Centered Design/Agile Development Process

MacroSys has adopted an iterative design and development process based on User Centered Design (UCD) and Agile principles (*https://www.agilealliance.org/agile101/12-principles-behind-the-agile-manifesto/*). This approach involves reaching out to the user community to understand their needs for a new CTPP data access solution, and then iteratively designing, developing, and testing the solution until maximum user satisfaction is achieved. Unlike the traditional waterfall project development process (*https://www.lucidchart.com/blog/waterfall-project-management-methodology*), the Agile process allows for ongoing user feedback throughout the design and development cycle, which can better accommodate changing user needs and provide a deeper understanding of the progress of development.

Guided by these design and development principles, CTPP users across the country participated in both the Initial Phase I and Initial Phase II Design Meetings. The users were invited to answer questions regarding their basic CTPP data needs, such as what data they need from the CTPP, when they need it, where they need data coverage, and why they need the data. The key takeaways from these meetings provided guidance to the contractor in order to map out the user landscape and know which user stories to prioritize and how to refine existing user stories as they develop the new CTPP data access solution.

Building on this effort, the CTPP user community, including those who participated in initial design meetings and new testing recruits, will provide their feedback on the 'first draft' GUI design concepts during the beta testing phase. Completing the first Beta Testing cycle, volunteer testers will put the first prototype GUI design through its paces, reviewing the search logic

It is not too late to volunteer if you would like to participate as a Beta Tester. You can sign up via the following link: *https://forms.office.com/r/fBPwnrnG62*

'usability', looking for bugs, etc., providing feedback to the design team to make improvements. The next cycle of Beta Tests will include new upgraded design GUI data search capabilities, including the previously suggested improvements, and continue to iterate Beta Test cycles until the final CTPP GUI design is achieved.

Big Data Technology and Column-oriented vs Row-based Data Store

MacroSys has integrated big data technology as a core component of the new data access solution. They specifically looked into using column-oriented data engines, such as Apache Hadoop and Apache Spark, to store CTPP tabulations, both historical datasets (around 350GB) and future releases. These technologies aim to provide fast response times, scalability, high availability, and fault tolerance by utilizing distributed processing and storage.

By contrast, traditional relational database management systems such as PostgreSQL/PostGIS (*https://www.postgresql.org/*) offer advanced geospatial capabilities and ease of integration with other GIS systems. Therefore, this solution can be used to store geospatial data associated with the CTPP, since these are relatively small files (less than 10GB). The goal is to create a system that combines the benefits of both big data technology and advanced geospatial functionality.

CTPP Data API and Hackathon

While a graphical user interface (GUI) is designed for human interaction with a web application, an Application Program Interface (API) allows other applications to interact with databases via code. As part of the re-build of the CTPP data access toolkit, a CTPP Data API has been created allowing users to write scripts that obtain CTPP data using programming languages such as Python, R, or JavaScript. Unlike desktop software APIs such as ESRI's ArcObjects, web services APIs allow data requests to be made over the Internet using HTTP protocols. This allows the applications communicating with each other to be totally agnostic as to each other's platforms, operating systems, or location on-premises versus in a computing cloud.

Unlike a graphical user interface which supports a limited set of pre-identified workflows and actions, an API can be used to create applications that incorporate other software and data, which means that the potential use cases are more open-ended but difficult to identify in advance. This is why Manhan organized a "hackathon" in which early access to the API would be provided to potential users who compete to build their own applications using it. Such competitions have become commonplace in software development and "data for good" communities, and some transportation planning agencies have organized hackathons in order to explore potential uses of new data sets or solicit creative solutions to difficult problems.

Hackathon participants were recruited via the TMIP and CTPP listservs, via word of mouth to various technical and software development communities, and from contract lists built up during previous outreach conducted by Manhan. A total of 27 people signed up to participate in the hackathon. Manhan worked with MacroSys to generate and distribute API keys to all participants as well as developing a documentation and demonstration page for the API, along with example scripts in Python, R and JavaScript to illustrate basic usage.

Participating teams were given about a month to develop their applications. As anticipated given the "early access" nature of the trial, some teams did encounter and report issues with the draft API, which were quickly fixed by MacroSys (within 1-2 business days). Early in the hackathon period, a webinar was provided, giving a chance for participants to interact directly with the team of independent judges selected by the Director of AASHTO's CTPP Program as well as the Chair of the CTPP Data Access Subcommittee.

The top-ranked entry, submitted by Suzanne Childress from the Puget Sound Regional Council (PSRC), paired a very useful web-based data visualization app (pictured below) with an R library for accessing the CTPP data via the new API (see their GitHub site at *https://github.com/psrc/psrcctpp* for details). Suzanne said that the hackathon was "a very fun opportunity for our team to work on a useful product, while building skills." *Figure 1 is a screenshot of the PSRC site that allows the user to map select CTPP tables, such as means of transportation, at the workplace or residential location. This image depicts the mode share of workers arriving via single-occupant vehicle.*
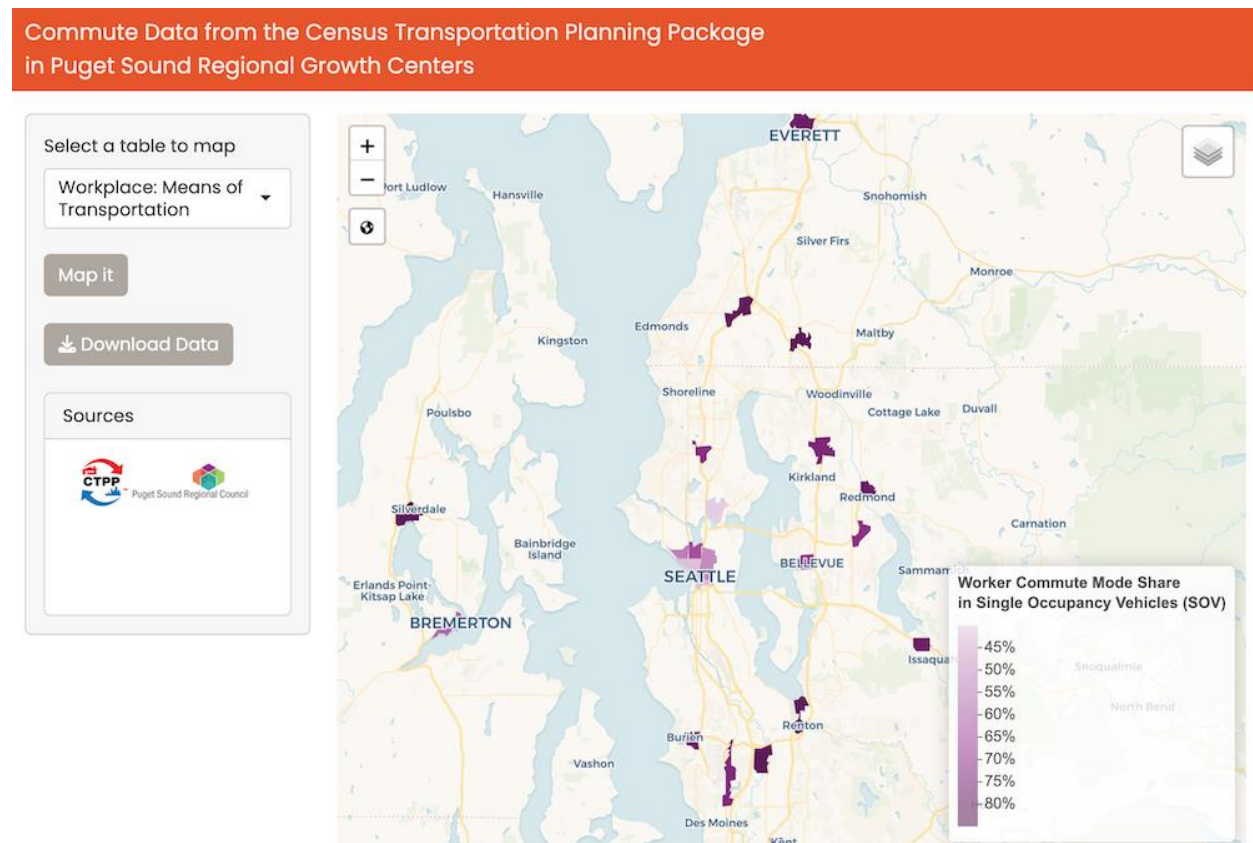

Figure 1. Top-ranked entry from Suzanne Childress and the Puget Sound Regional Council (PSRC)
https://www.psrc.org/

Second place went to Steve Hossack and Mark Jackman, from the Orange County Transportation Authority, who developed an ArcGIS web app for visualizing distribution of commute origin TAZs by destination TAZ (or vice versa). This project also included a Python geoprocessing script for pulling data. *Figure 2 is a screenshot of the OCTA site that allows the user to select origin and destination zones for depicting worker flows. The zone outlined in green is the origin zone and the darker the fill in the other zones the more trip destinations are there.*
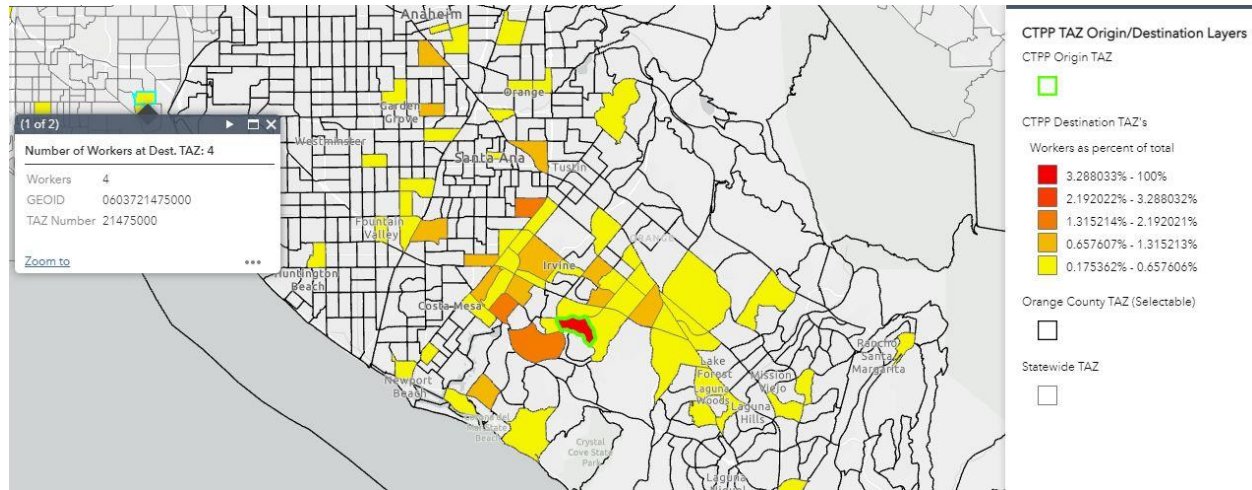

Figure 2. Second place entry from Steve Hossack and Mark Jackman at the Orange County Transportation Authority *https://www.octa.net/*

Finally, third place went to Zihao Jin, from Boston's Central Transportation Planning Staff, who showcased two data explorers developed explicitly for the freshly released data API so-called "Across State" and "Stack State", developed using the Matrix Platform with Python scripting.

Next Steps: Beta Testing the Updated Web UI

In February 2023, MacroSys is anticipated to deliver the updated web-based GUI for accessing CTPP data to AASHTO for internal (i.e., Manhan staff) testing. Soon thereafter, the GUI will be made available to CTPP users for Beta testing, so that MacroSys can gather user feedback and make revisions in order to ensure the system's usability. Additional Beta test rounds will continue throughout 2023. To stay updated on the project's progress, subscribe to the CTPP listserv (*https://listserv.transportation.org/mailman3/lists/ctpp.listserv.transportation.org/*) and check Manhan's CTPP update project page (*https://manhan.co/CTPP-Update.html*) regularly.

CTPP Contact List

AASHTO
Penelope Weinberger
Phone: (202) 624-3556
Email: *pweinberger@aashto.org*

Sondra Rosenberg, NVDOT
Chair, CTPP Oversight Board
Phone: (775)-888-7440
Email: *srosenberg@dot.nv.gov*

Guy Rousseau, Atlanta Regional Commission
Vice Chair, CTPP Oversight Board
Phone: (404) 463-3274
Email: *GRousseau@atlantaregional.com*

U.S. Census Bureau: Journey to Work and Migration Statistics Branch
Brian McKenzie
Phone: (301) 763-6532
Email: *brian.mckenzie@census.gov*

Federal Transit Administration (FTA)
Jeffrey Roux
Phone: (202) 366-4043
Email: *jeffrey.roux@dot.gov*

Bureau of Transportation Statistics (BTS)
Clara Reschovsky
Phone: (202) 366-2857
Email: *clara.reschovsky@dot.gov*

Federal Highway Administration (FHWA)
Joseph Hausman
Phone: (202) 366-9629
Email: *Joseph.Hausman@dot.gov*

TRB Committees
Michael Fontaine, Virginia DOT
Chair, TRB Urban Data Committee
Email: *michael.fontaine@vdot.virginia.gov*

Kathleen Yu
North Central Texas Council of Governments
TRB Census Subcommittee Co-Chair
Phone: (817) 608-2343
Email: *kyu@nctcog.org*

CTPP Technical Support, News Editor
Rob Schiffer, AICP
Phone: (850) 570-8958
Email: *rob@metroanalytics.com*

2012-2016 CTPP Data: *https://ctpp.transportation.org/2012-2016-5-year-ctpp/*
FHWA CTPP website: *https://www.fhwa.dot.gov/planning/census_issues/ctpp/*
FHWA website for Census issues: *https://www.fhwa.dot.gov/planning/census_issues*
AASHTO website for CTPP: *https://ctpp.transportation.org*
1990 and 2000 CTPP data downloadable via Transtats: *https://transtats.bts.gov/*
TRB Subcommittee on census data: *http://www.trbcensus.com*

CTPP Listserv

The CTPP Listserv serves as a web-forum for posting questions and sharing information on Census and ACS. Currently, more than 700 users are subscribed to the listserv. To subscribe, please register by completing a form posted at:
https://listserv.transportation.org/mailman3/lists/ctpp.listserv.transportation.org/

On the form, you can indicate if you want emails to be batched in a daily digest. The website also includes an archive of past emails posted to the listserv.