"THE PUMS & IPUMS OF THE 2000 CENSUS"

EDREECE "ED" AZIMI
VIRGINIA DEPARTMENT OF TRANSPORTATION
Northern Virginia District Transportation Planning
TRB Census Data Conference, May 11-14, 2005, Irvine, CA

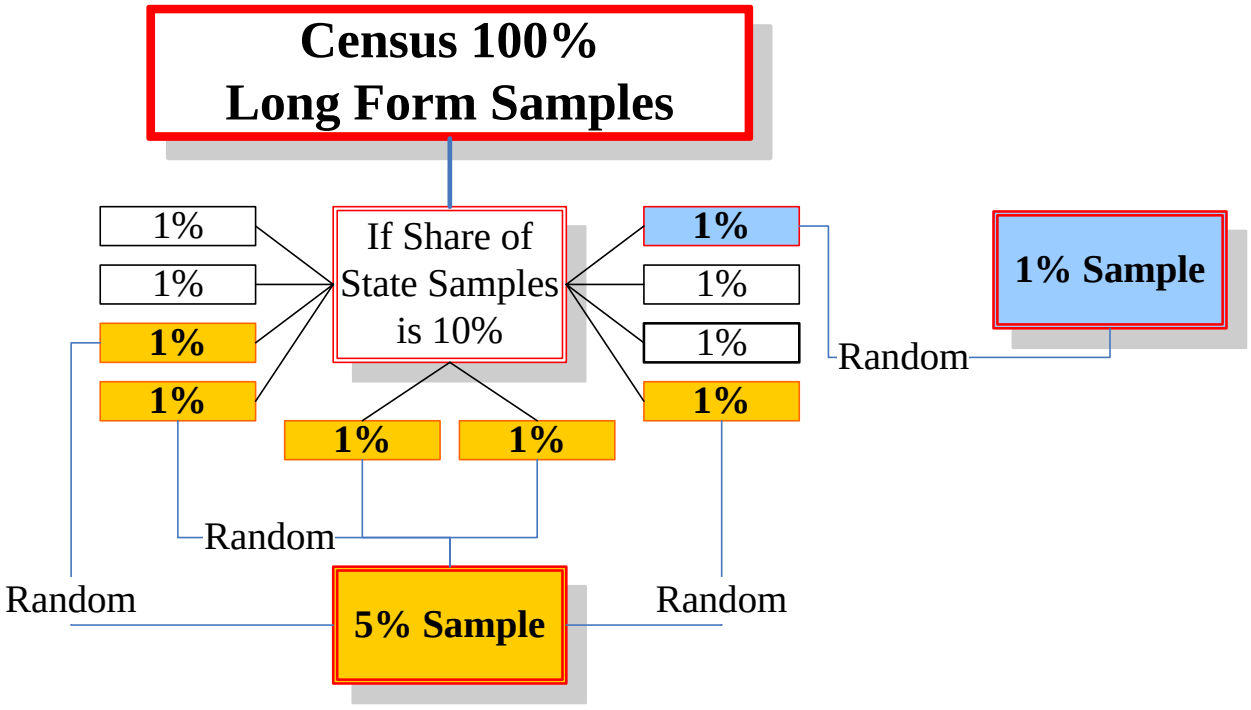
Abstract (#CD-0037)

To summarize and analyze carpoolers by origin, destination, job type, and corridor, our most important data was the 2000 Census Transportation Planning Package (CTPP). However, before each of the three (3) parts of the CTPP was available, we had to use the 2000 Census data (the original source of CTPP). Then we learned about the Public Use Microdata Sample (PUMS) and Integrated Public Use Microdata Series (IPUMS) as alternative sources for transportation data. In this paper we are sharing the experience and the significant uses of these two sources of data – when to use the PUMS versus the IPUMS. This paper also explains the difficulty we experienced employing these two sources of records. For example, the "Should or shouldn't" use of 5% sample vs. the 1% sample. Overall, this paper will provide the pros and cons of the PUMS and IPUMS.

Introduction

The *Carpooling Characteristics* is a State Planning & Research (SPR) project, supporting the Washington region, will summarize and analyze carpoolers by origin, destination, job type, and corridor used. It requires input from the 2000 Census Transportation Planning Package (CTPP) and the local MPO's version of that package. Currently, the MPO (MWCOG) is processing the 2000 Census data and converting it to a TAZ (Transportation Analysis Zone) format that is compatible with the 1990 CTPP format. In the absence of CTPP Part 3 data, we looked for the sources of data in the Census database to accommodate our needs for the *Journey-to-work* information. As a result, we found about the PUMS database and later, gladly, we learned about the IPUMS as a giant data source. The NOVA jurisdictions for study are: Arlington Co., Fairfax Co., Fauquier Co., Loudoun Co., Prince William Co., Spotsylvania Co., Stafford Co., City of Alexandria, City of Fairfax, City of Falls Church, City of Fredericksburg, City of Manassas and City of Manassas Park.

1% vs. 5% Sample Selection



2000 Census Washington Metro Area PUMA

Highlights

- ◊ No. of Jurs. with single PUMA = 2.
- ◊ No. of jurs. with multiple PUMA = 2.
- ◊ One case of 4 jurs. with single PUMAs with 3 unwanted jurs.
- ◊ Two cases of 2 jurs. with single PUMAs with 1 unwanted jurs.

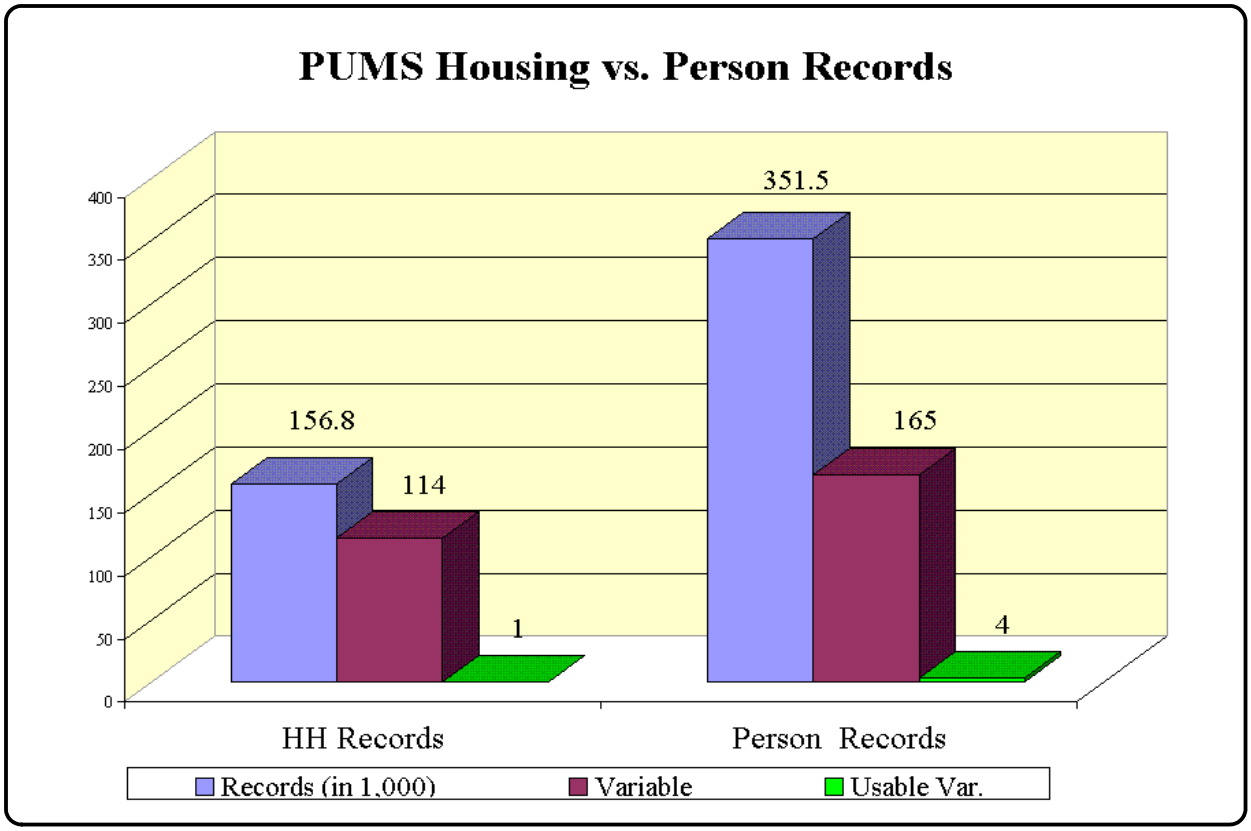
PUMS Variables & Records

Items in the housing record include:

- Bedrooms
- Condominium status
- Contract rent (monthly rent)
- Cost of utilities and fuels
- Family income
- Family, subfamily, and household relationships
- Farm status and value
- Fire, hazard, and flood insurance
- Food Stamps recipiency
- Fuels used
- Gross rent
- House heating fuel
- Household income
- Household type
- Kitchen facilities
- Linguistic isolation*
- Meals included in rent
- Mortgage status and selected monthly owner costs
- Plumbing facilities
- Presence and age of own children
- Presence of subfamilies in household
- Property value
- Real estate taxes
- Residence State
- Rooms
- Sewage disposal
- Source of water
- Telephone in housing unit
- Tenure
- Units in structure
- Vacancy status
- Vehicles available
- Year householder moved into unit
- Year structure built

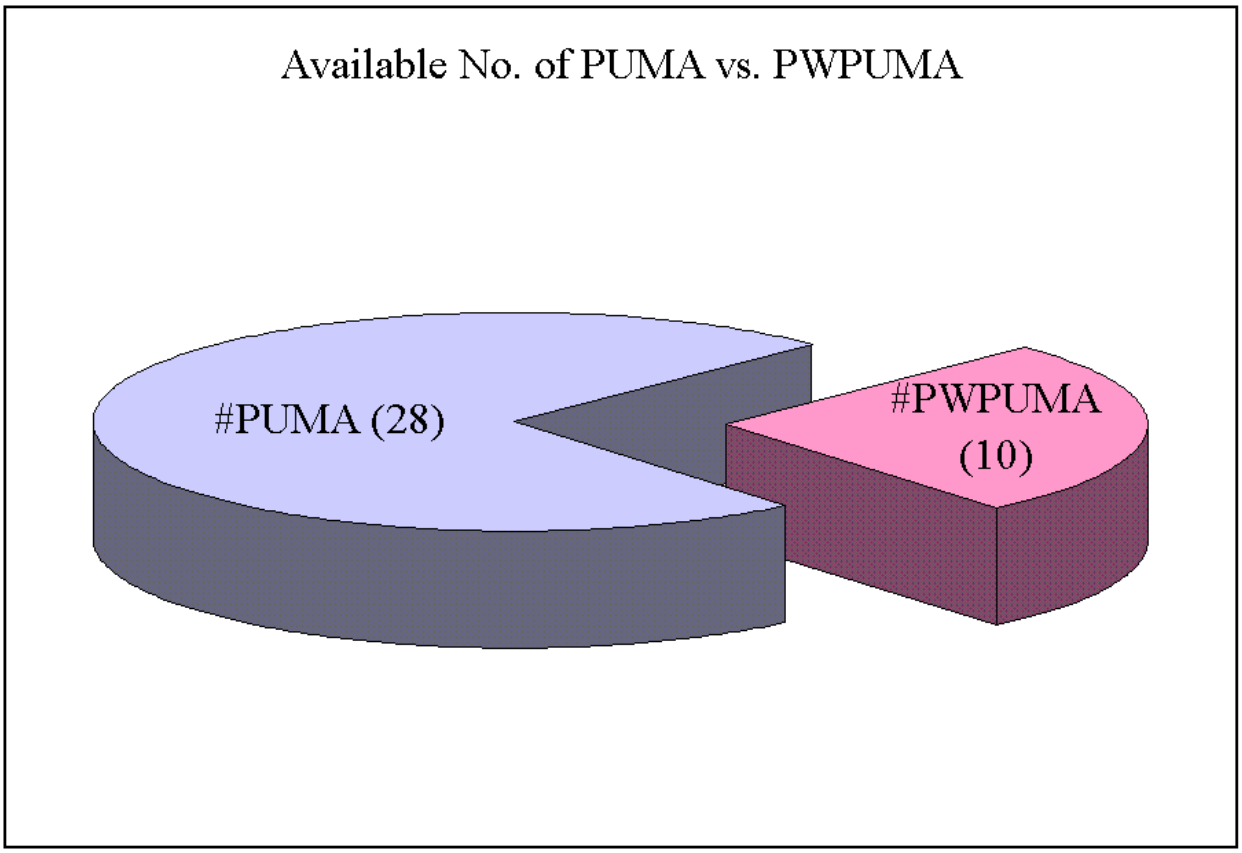
Items in the person record include:

- Ability to speak English
- Age
- Ancestry
- Citizenship
- Class of worker
- Disability status
- Educational attainment
- Fertility
- Hispanic origin
- Hours worked
- Income by type
- Industry
- Language spoken at home
- Last week work status
- Marital status
- Means of transportation to work
- Migration
- Military status, periods of active duty military service, veteran period of service
- Mobility status
- Occupation
- Personal care limitation
- Place of birth
- Place of work
- Poverty status
- Race
- Relationship
- School enrollment and type of school
- Sex
- Time of departure for work
- Travel time to work
- Vehicle occupancy
- Weeks worked
- Work status
- Work limitation status
- Year of entry


PUMS Housing vs. Person Records
400
350
300
250
200
150
100
50
0
156.8
114
1
351.5
165
4
HH Records
Person Records
Records (in 1,000)
Variable
Usable Var.

Relationship Between Place-of-Residence PUMA & Place-of-Work (PW) PUMA

State Code (STATEFIP)	State Name	County/City	Place of Residence PUMA	Place of work PUMA (PWPUMA00)
11	District of Columbia	D.C.	00101-00105	100
24	Maryland	Montgomery	01101-01107	1100
		Prince Georges	01201-01204	1200
51	Virginia	Arlington	100	100
		Alexandria	200	200
		Fairfax	00301-00305	300
		Prince William	00501-00502	500
		Loudoun, Fauquier, Clark, Warren	600	600
		Stafford, King Georges	800	800
		Spotsylvania, Culpeper	900	900


Available No. of PUMA vs. PWPUMA
#PUMA (28)
#PWPUMA (10)

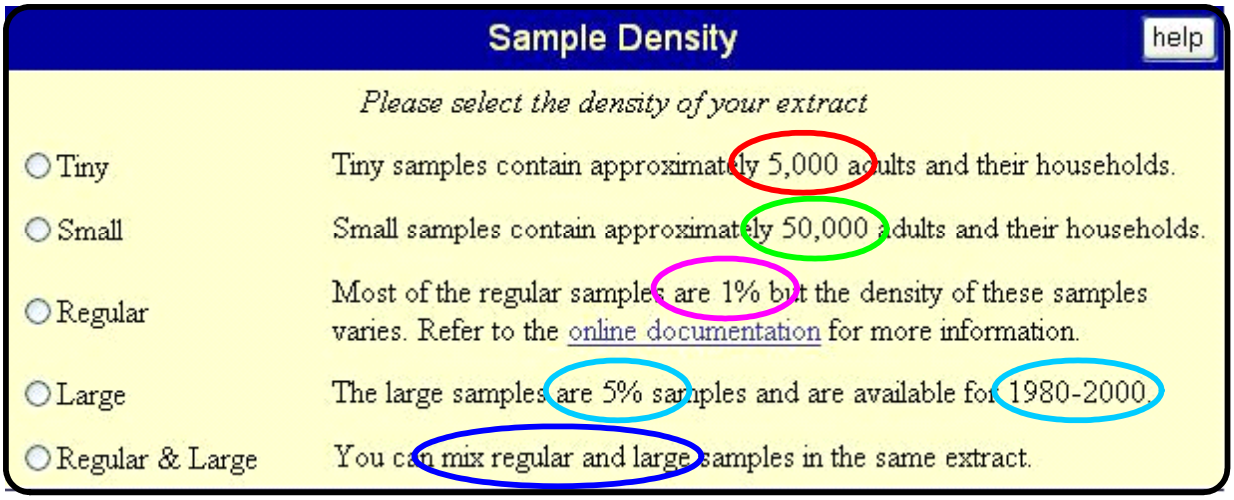
IPUMS Sample of Variables & Sample Density

Appliances, Mechanical, Other Variables (Household)

Variable Name	Variable Description	Variable Description	Codes and Frequencies	1850	1860	1870	1880	1900	1910	1920	1940	1950	1960	1970	1980	1990	2000
PHONE	Telephone availability	Desc	Codes	.	.	.	.	.	.	.	.	.	X	X	X	X	X
TV	Television set	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F1	.	.	.
UHF	UHF-equipped television set	Desc	Codes	.	.	.	.	.	.	.	.	.	.	F1	.	.	.
RADIO	Radio	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F1	.	.	.
WASHER	Clothes washing machine	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F1	.	.	.
DRYER	Clothes dryer	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F1	.	.	.
DISHWSHR	Dishwasher	Desc	Codes	.	.	.	.	.	.	.	.	.	.	F1	.	.	.
FREEZER	Home food freezer	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F1	.	.	.
AIRCON	Air conditioning	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F2	X	.	.
HEATING	Heating equipment	Desc	Codes	.	.	.	.	.	.	.	.	.	X	X	X	.	.
FUELCOOK	Cooking fuel	Desc	Codes	.	.	.	.	.	.	.	.	.	.	F1	X	.	.
FUELHEAT	Home heating fuel	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F1	X	X	X
FUELWATR	Water heating fuel	Desc	Codes	.	.	.	.	.	.	.	.	.	.	F1	X	.	.
AUTOS	Automobiles available	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F2	X	.	.
TRUCKS	Trucks and vans available	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	X	.	.
VEHICLES	Vehicles available	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	.	X	X
LINGISOL	Linguistic isolation	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	.	X	X

Place of Work and Travel Time Variables (Person)

Variable Name	Variable Description	Variable Description	Codes and Frequencies	1850	1860	1870	1880	1900	1910	1920	1940	1950	1960	1970	1980	1990	2000
PWSTATE1	Place of work: state, 1960 and 1970	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F2	.	.	.
PWSTATE2	Place of work: state, 1980 to 2000	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	X	X	X
PWMETRO	Place of work: metropolitan area	Desc		.	.	.	.	.	.	.	.	.	.	.	S,M	X	1,5
PWCITY	Place of work: city	Desc		.	.	.	.	.	.	.	.	.	.	.	S,M	X	1,5
PWTYPE	Place of work: metropolitan status	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F2	S,M	M	1,5
PWTYPE00	Place of work: metropolitan status, 2000	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	.	.	1,5
PWURBTYP	Place of work: urbanized area	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	UR	.	.
PWSIZE	Place of work: place size	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	UR	.	.
PWCNTYGP	Place of work: county group	Desc		.	.	.	.	.	.	.	.	.	.	.	S,M	.	.
PWPUMA	Place of work: PUMA	Desc		.	.	.	.	.	.	.	.	.	.	.	.	X	.
PWPUMA00	Place of work: PUMA, 2000	Desc		.	.	.	.	.	.	.	.	.	.	.	.	.	5
PWPUMAS	Place of work: Super-PUMA, 2000	Desc		.	.	.	.	.	.	.	.	.	.	.	.	.	1,5
PWMET98	Place of work: metropolitan status, 1980	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	S,M	.	.
TRANWORK	Means of transportation to work	Desc	Codes	.	.	.	.	.	.	.	.	.	X	F2	X	X	X
CARPOOL	Carpooling	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	X	X	X
RIDERS	Vehicle occupancy	Desc	Codes	.	.	.	.	.	.	.	.	.	.	.	X	X	X
TRANTIME	Travel time to work	Desc		.	.	.	.	.	.	.	.	.	.	.	X	X	X
DEPARTS	Time of departure for work	Desc		.	.	.	.	.	.	.	.	.	.	.	.	X	X


Sample Density
help
Please select the density of your extract
Tiny
Tiny samples contain approximately 5,000 adults and their households.
Small
Small samples contain approximately 50,000 adults and their households.
Regular
Most of the regular samples are 1% but the density of these samples varies. Refer to the online documentation for more information.
Large
The large samples are 5% samples and are available for 1980-2000.
Regular & Large
You can mix regular and large samples in the same extract.

PUMS vs. IPUMS (Pros & Cons)

PUMS Records & Variables

- Housing Unit Record (VA)= 156,800, with 114 Variables (Use Var= 1, **Vehicles Available**)

- Person Records (VA)= 351,485, with 165 Variable (Use Var= 4, **Mean of Transportation to Work**, **Time Leaving for Work**, **Travel Time to Work**, and **Vehicle Occupancy**

Pros.

- Unique Identifier, A serial number that links these two records.
- Describes characteristics of individuals.

Cons.

- Records are Extremely Large.
- Data are codes.
- Requires Statistical or Report-writing Software.
- Extremely difficult to cross-tabulate Housing Unit Record and Person Record.
- The Census information describes characteristics of individuals.

What is IPUMS= Integrated Public Use Microdata Series, created at the University of Minnesota in October 1997. The IPUMS consists of twenty-seven high-precision samples, which span the censuses of 1850 to 2000 of the American population drawn from fourteen Federal Census.

Pros.

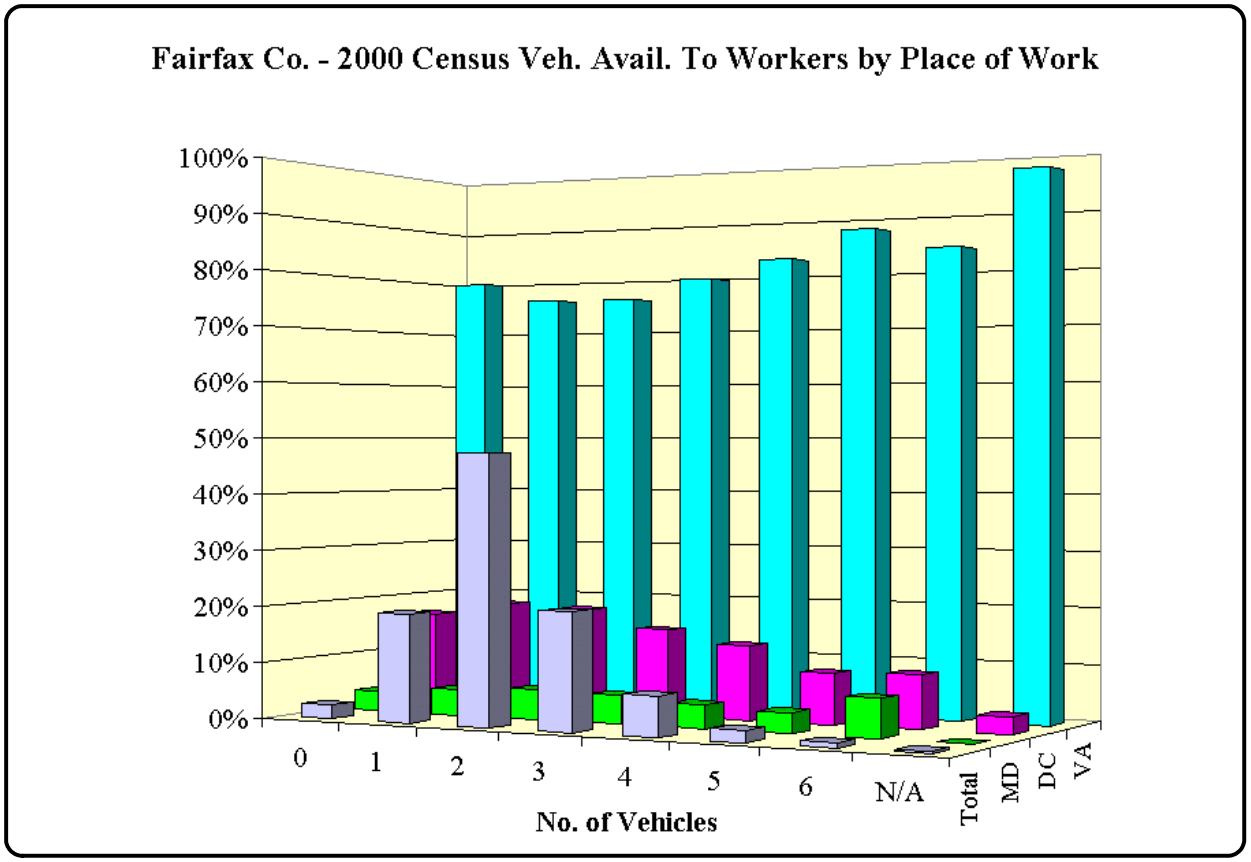
- The IPUMS is *microdata*, which means that it provides information about individual Persons and Households (HH).
- Individuals are sampled as parts of HH.
- All the IPUMS samples are **cluster samples**.
- Researchers are able to create tabulations tailored to their particular questions.
- Sample Density Selection.
- Samples span to 1850.
- Collectively comprise richest source of quantitative information on long-term changes in the American population.
- The IPUMS assigns uniform codes across all the samples.

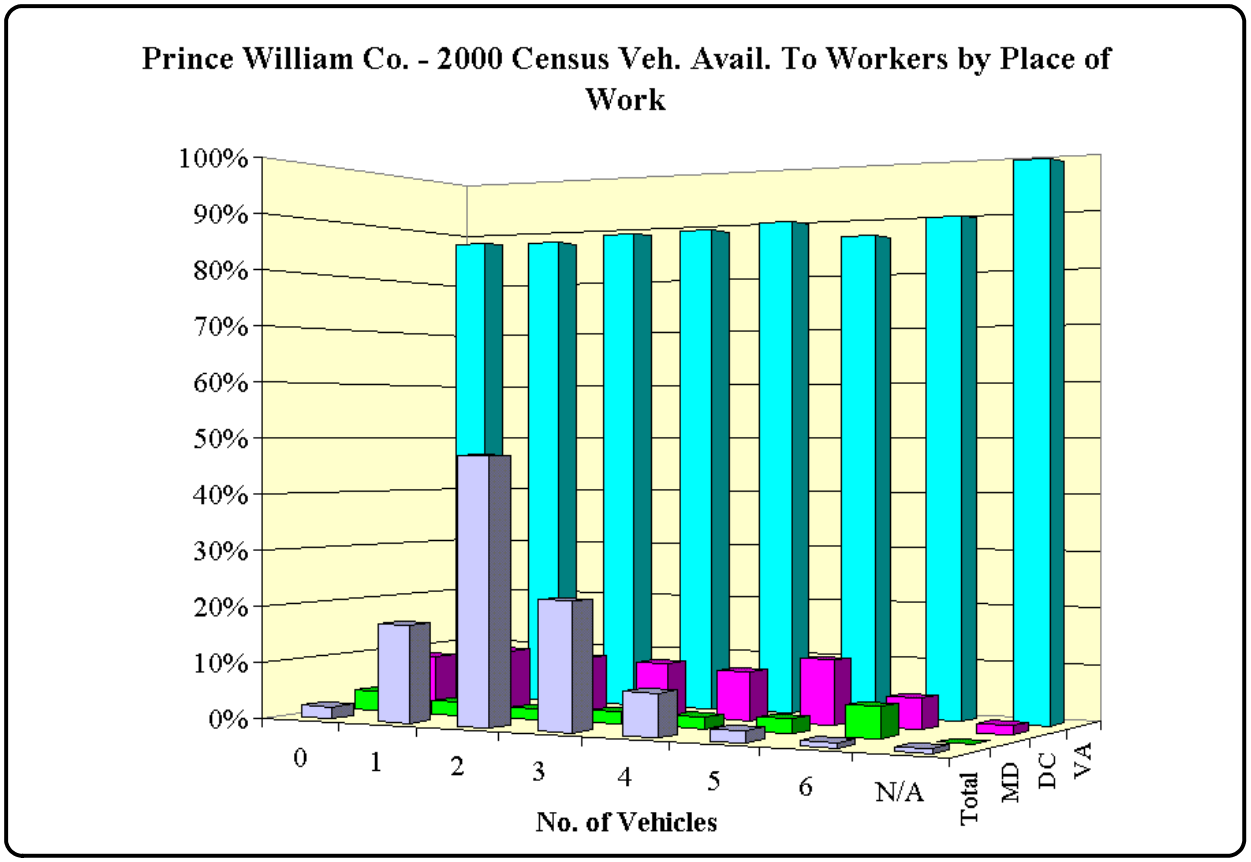
Cons.

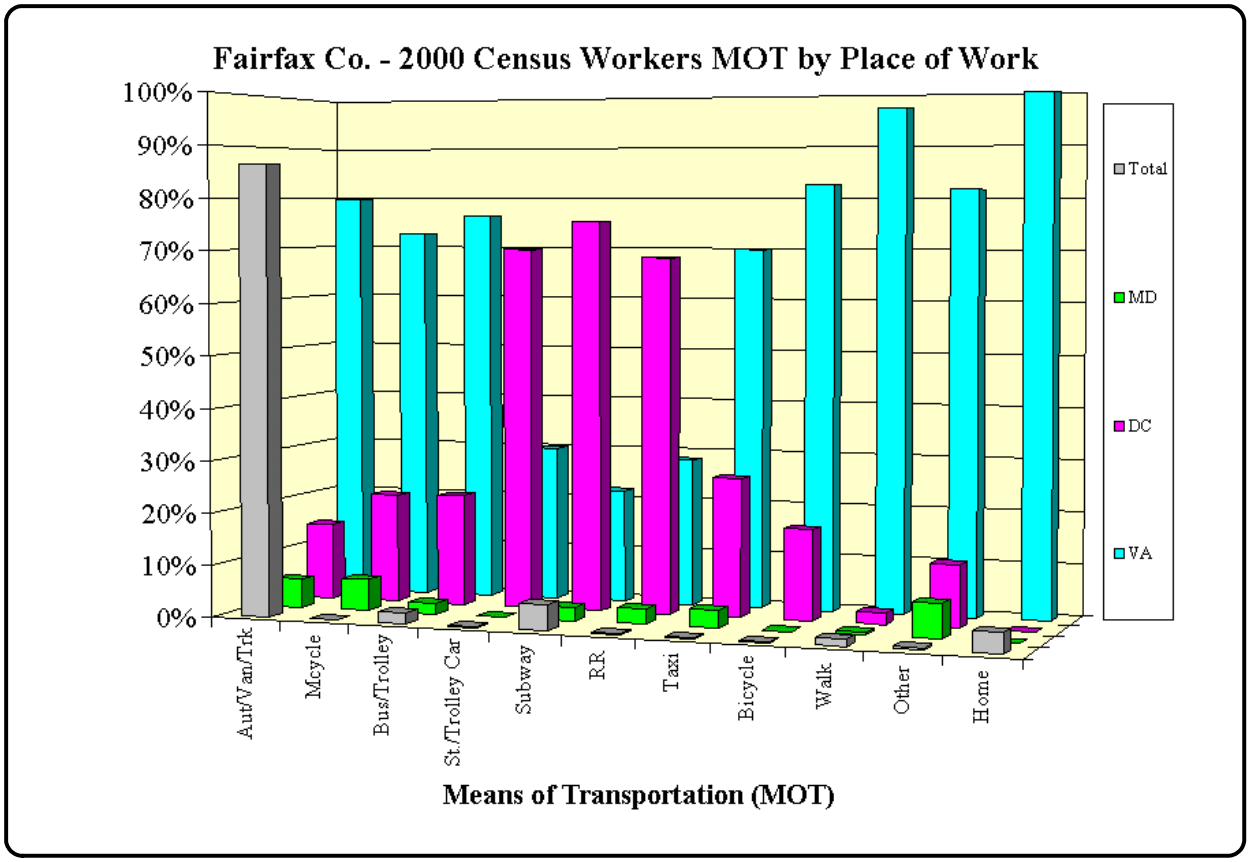
- Microdata do pose some limitations and are inappropriate for research that requires the identification of specific small geographic areas.

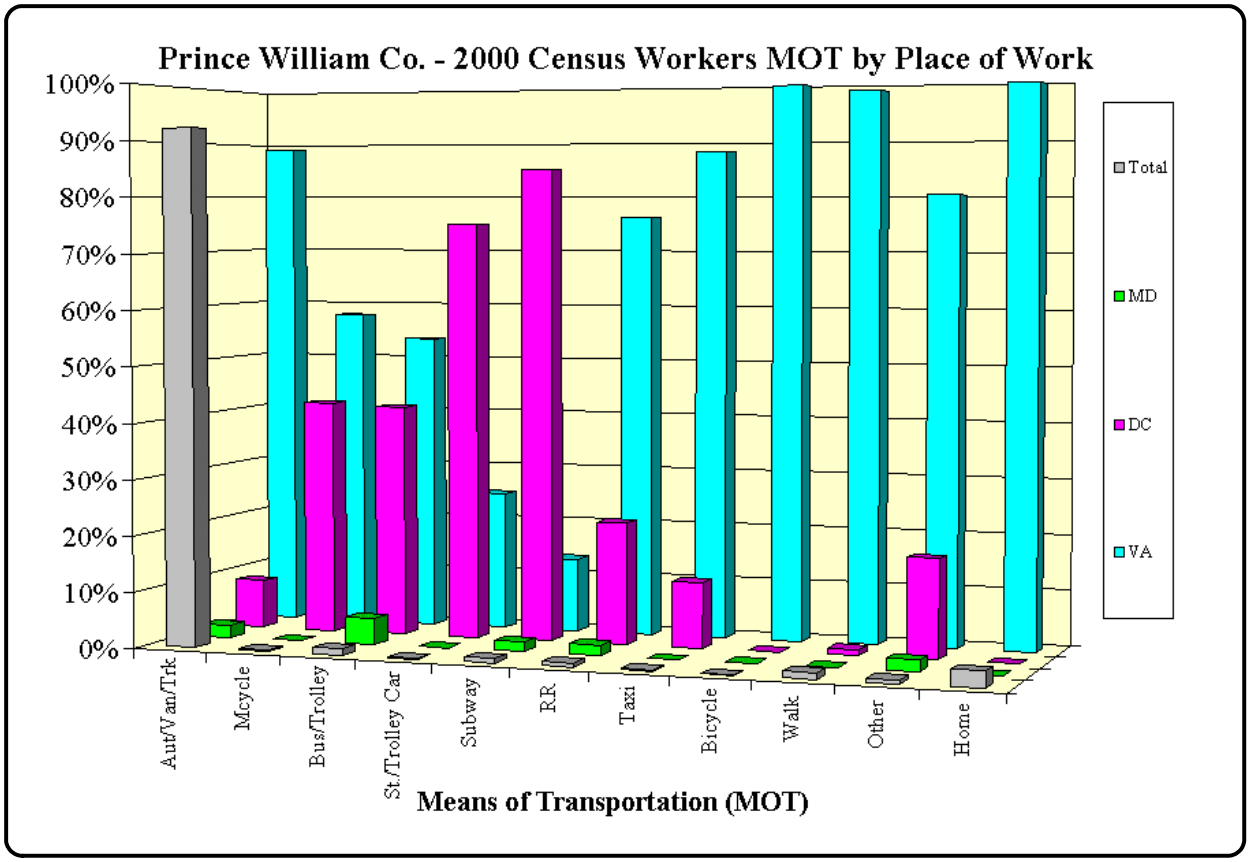
IPUMS 5% Sample Analysis Products (Only 4 out of 13 Jurisdictions' data were available)

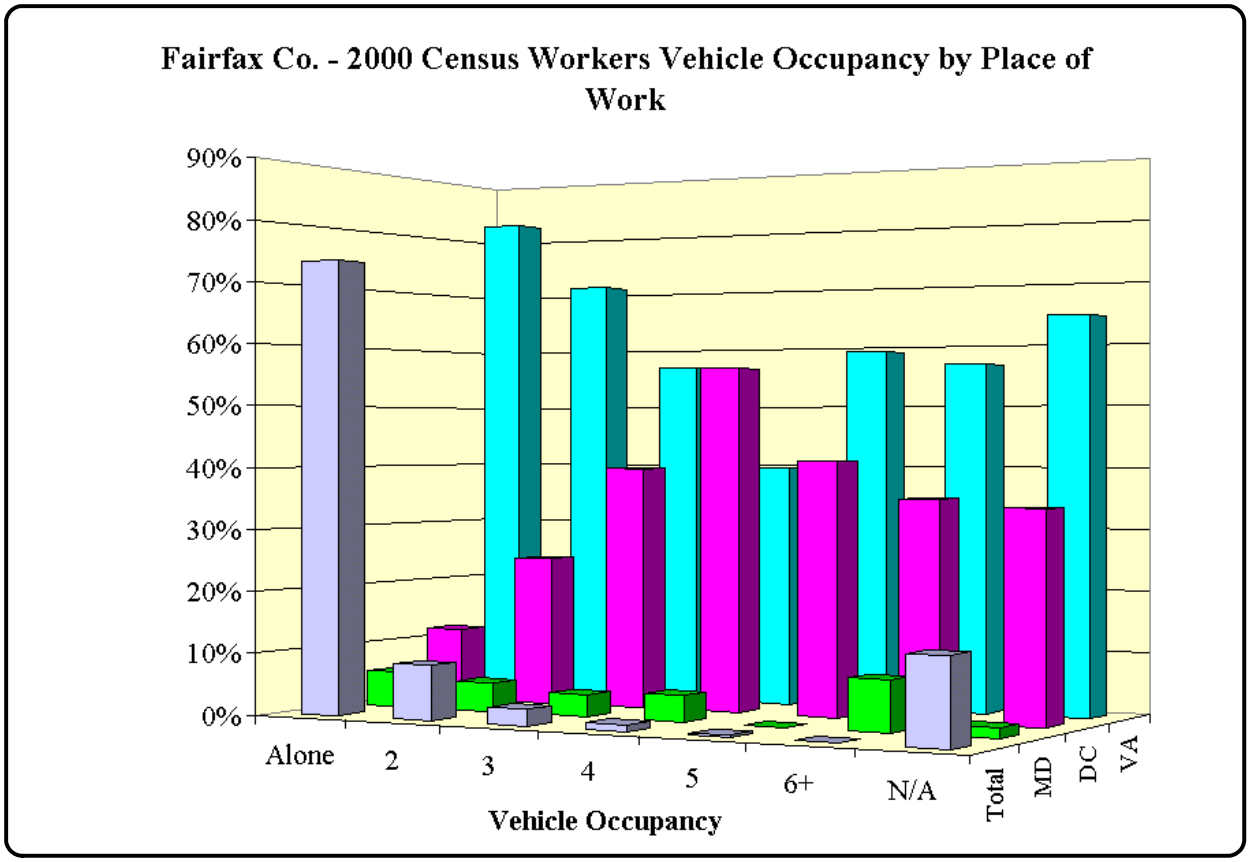
The Virginia jurisdictions under study for *Carpooling Characteristics* are part of Metropolitan Washington Council of Government "MWCOG" which embodies the National Capital Region Transportation Planning Board "TPB." (The TPB is the federally designated Metropolitan Planning Organization "MPO" for the region, and plays an important role as the regional forum for transportation planning.) The under study jurisdictions include seven (7) counties and Five (6) cities. The PUMS data are not available individually for all jurisdictions except for three Arlington, Fairfax and Prince William counties; and the City of Alexandria. The other four (4) counties are part of other grouped counties to make the PUMA. The Cities of Fairfax and Falls Church are part of Fairfax County. The Cities of Manassas and Manassas Park are part of Prince William County. Lastly, the City of Fredericksburg is part of Spotsylvania County. Therefore, the analysis of PUMS records were limited and must be combined into a group to depict the information for the place of work for "*Number of Vehicles*", "*Means of Transportation (MOT)*" and "*Vehicle Occupancy*". Samples of tabulated data are presented in graph for the two (2) of the three available individual counties and followed by figures of CTPP data including Loudoun County, which was not available by either PUMS & IPUMS data sources.

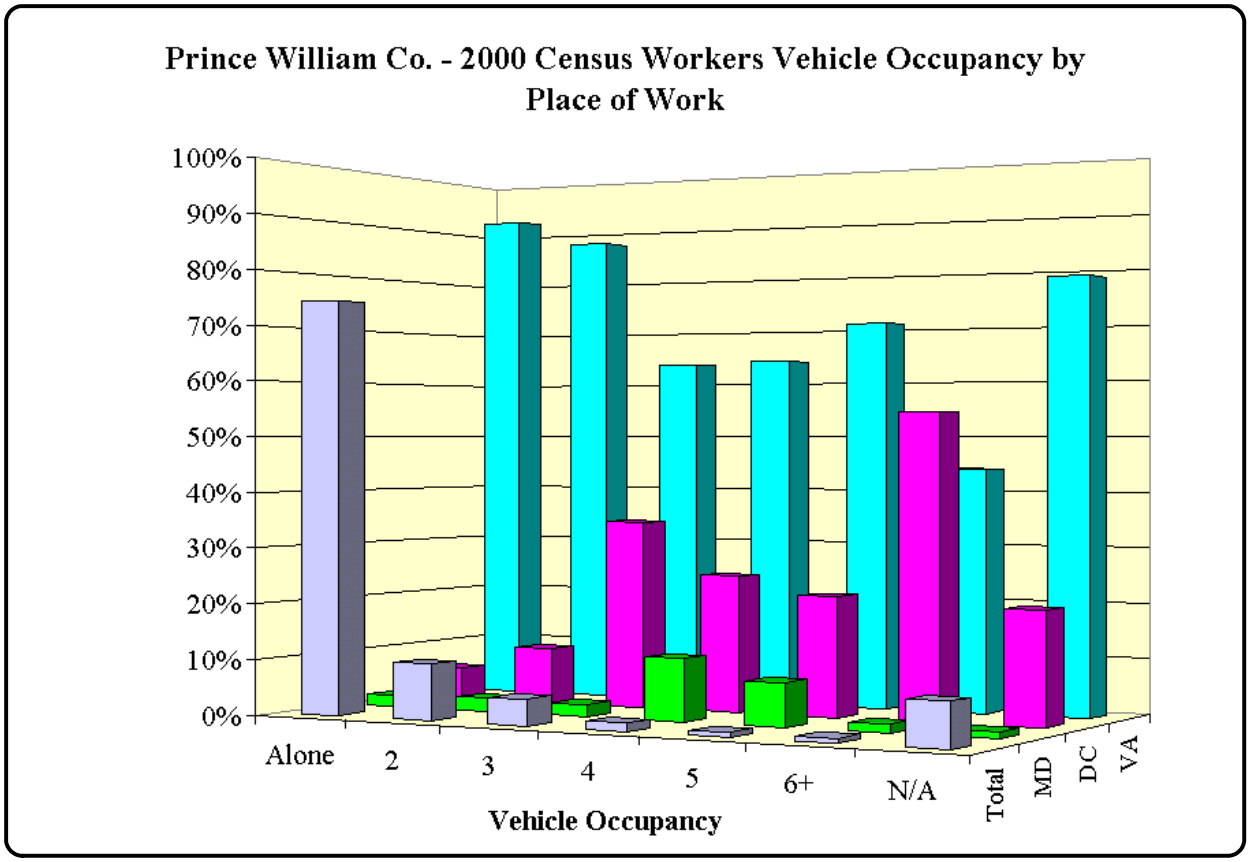

Fairfax Co. - 2000 Census Veh. Avail. To Workers by Place of Work
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
1
2
3
4
5
6
N/A
No. of Vehicles
Total
MD
DC
VA


Prince William Co. - 2000 Census Veh. Avail. To Workers by Place of Work
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
1
2
3
4
5
6
N/A
No. of Vehicles
Total
MD
DC
VA

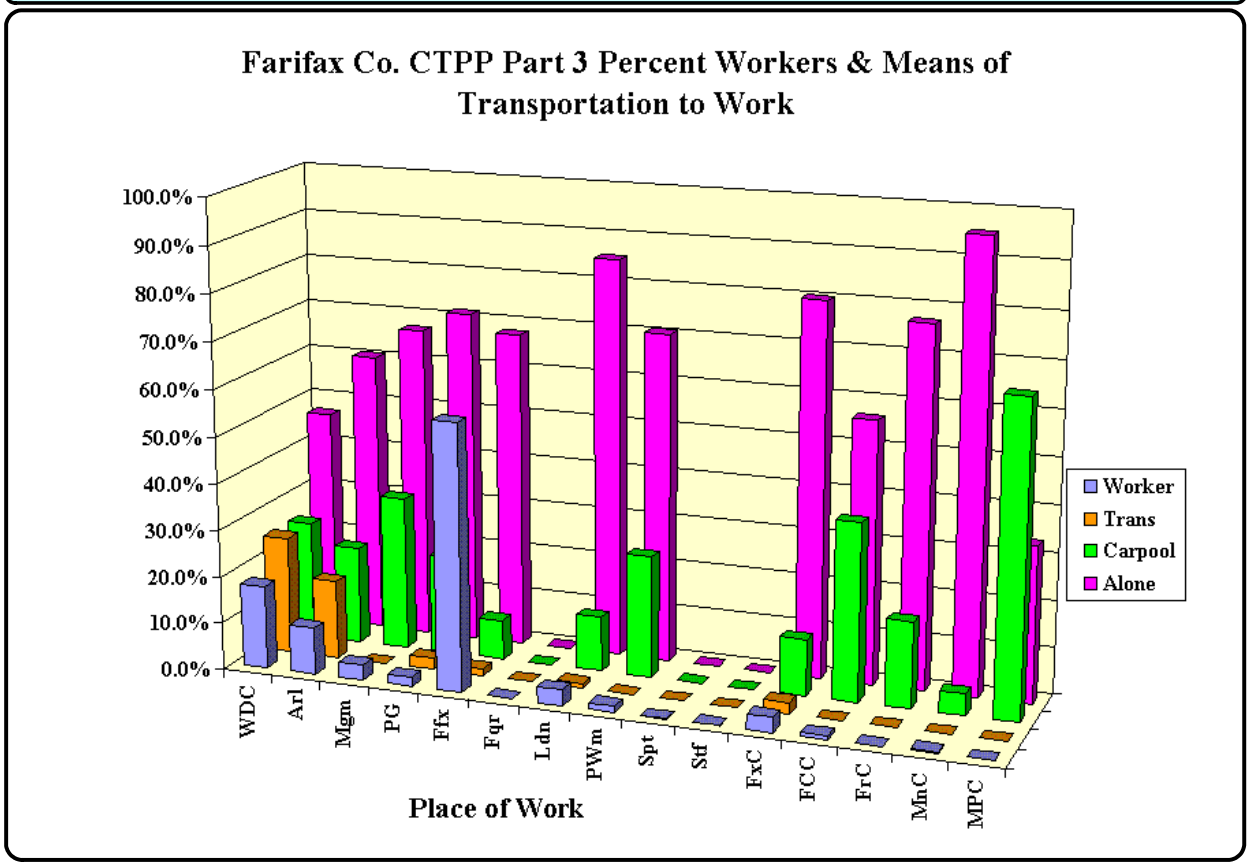

Fairfax Co. - 2000 Census Workers MOT by Place of Work
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Total
MD
DC
VA
Aut/Van/Trk
Mcycle
Bus/Trolley
St./Trolley Car
Subway
RR
Taxi
Bicycle
Walk
Other
Home
Means of Transportation (MOT)

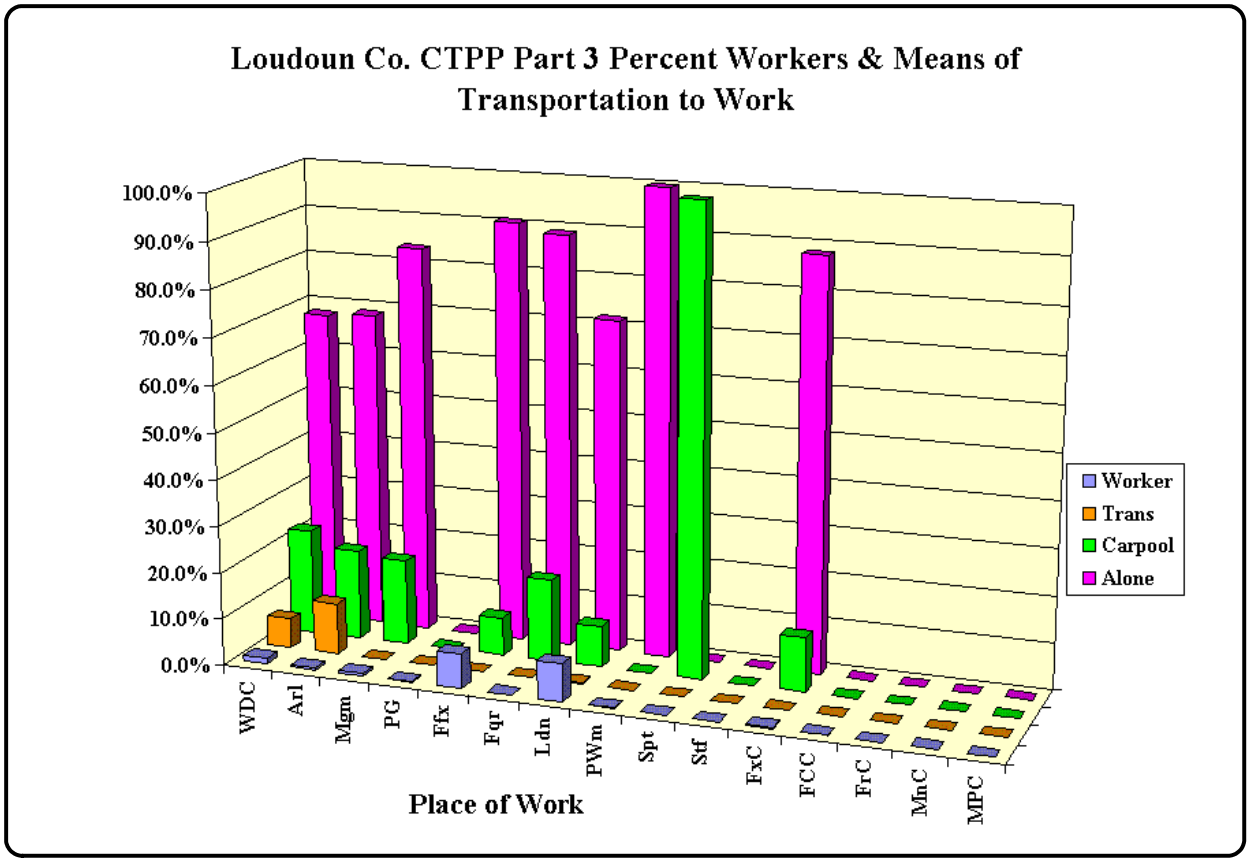

Prince William Co. - 2000 Census Workers MOT by Place of Work
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Total
MD
DC
VA
Aut/Van/Trk
Mcycle
Bus/Trolley
St./Trolley Car
Subway
RR
Taxi
Bicycle
Walk
Other
Home
Means of Transportation (MOT)

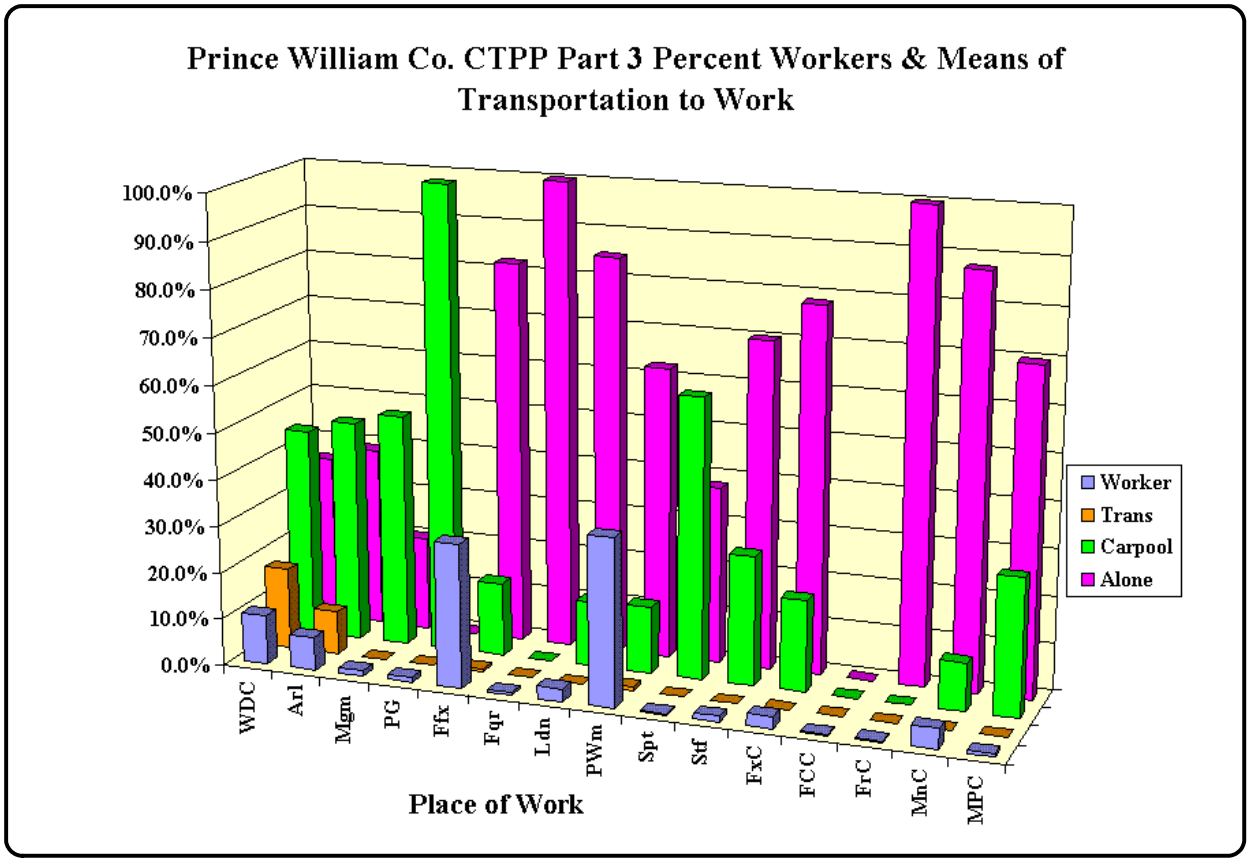

Fairfax Co. - 2000 Census Workers Vehicle Occupancy by Place of Work
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Alone
2
3
4
5
6+
N/A
Total
MD
DC
VA
Vehicle Occupancy

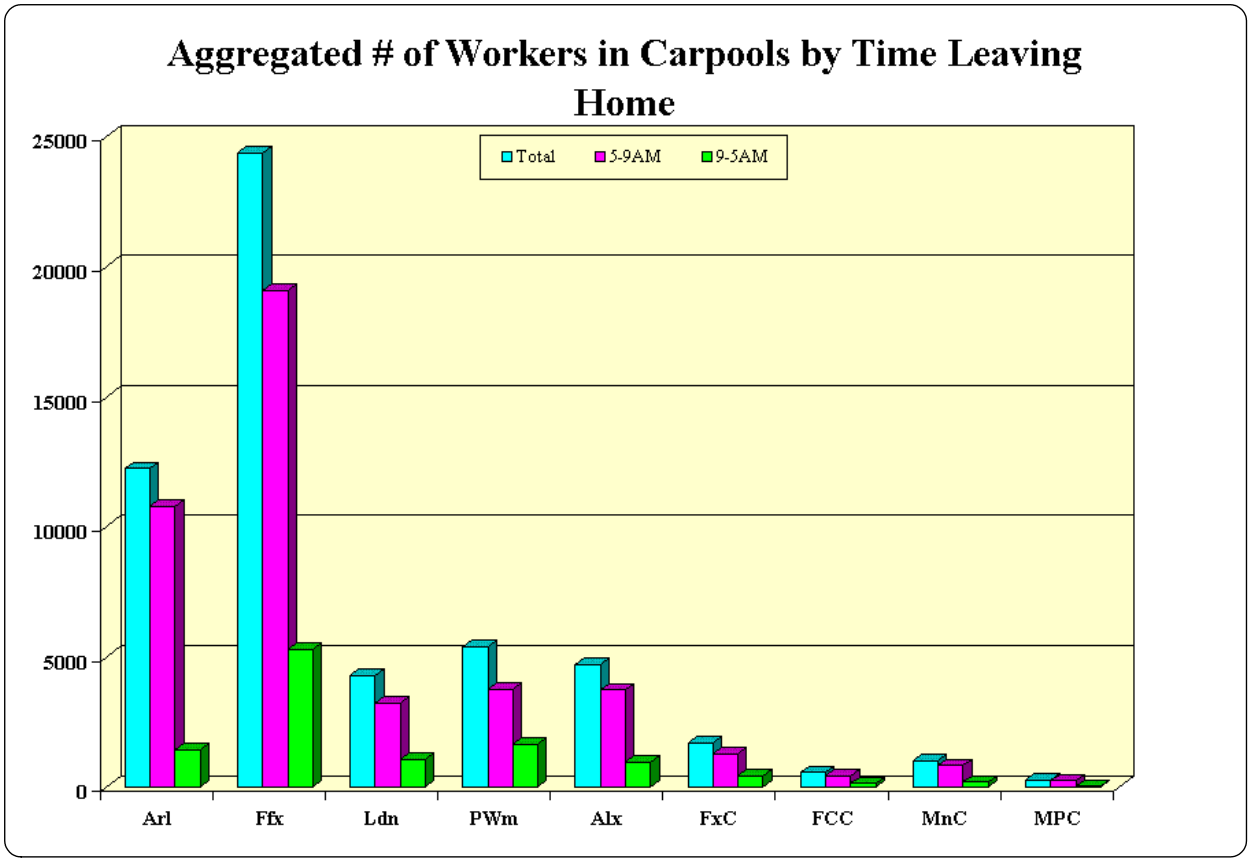

Prince William Co. - 2000 Census Workers Vehicle Occupancy by Place of Work
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Alone
2
3
4
5
6+
N/A
Total
MD
DC
VA
Vehicle Occupancy

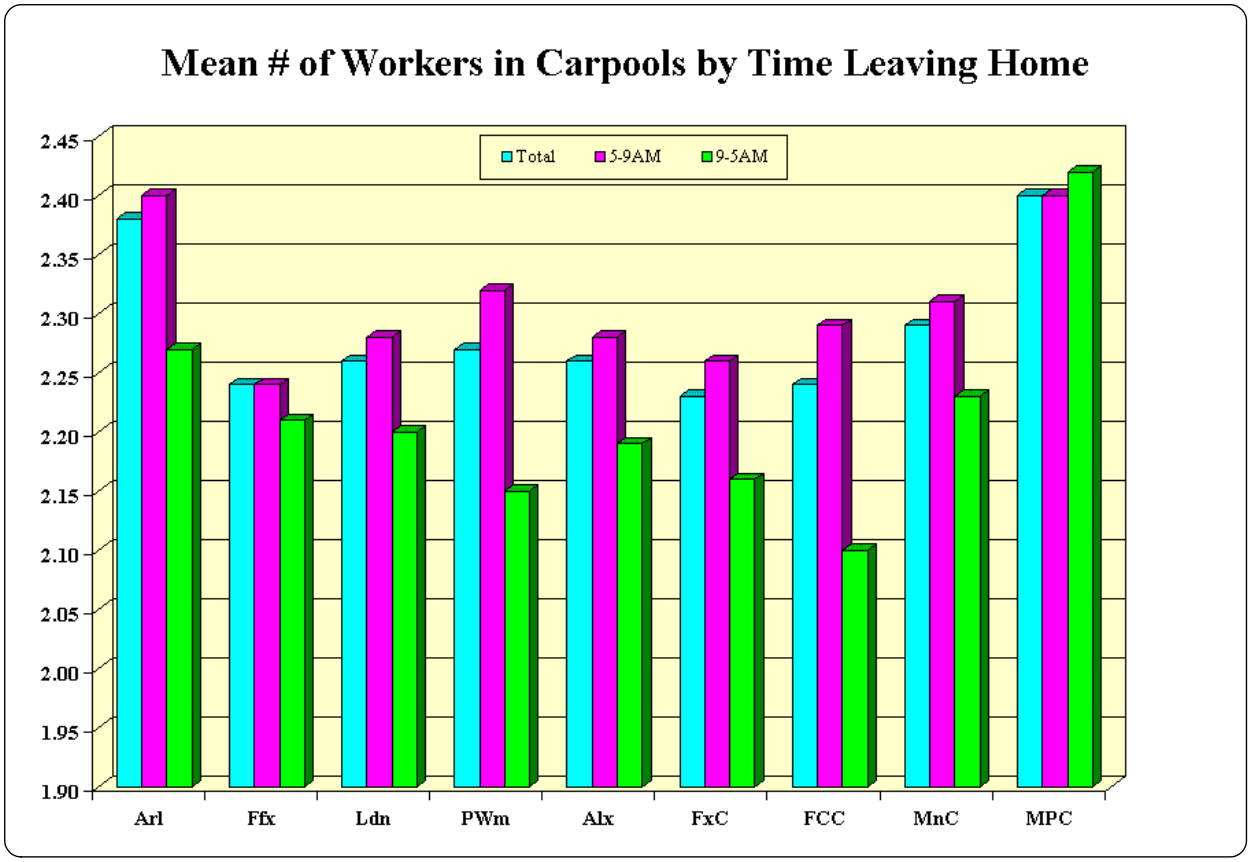
2000 Census CTPP Part 3 Products Sample (Data were available for all the jurisdictions in the study)




Loudoun Co. CTPP Part 3 Percent Workers & Means of Transportation to Work
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
WDC
Arl
Mgm
PG
Ffx
Fqr
Ldn
PWm
Spt
Stf
FxC
FCC
FrC
MnC
MPC
Place of Work
Worker
Trans
Carpool
Alone

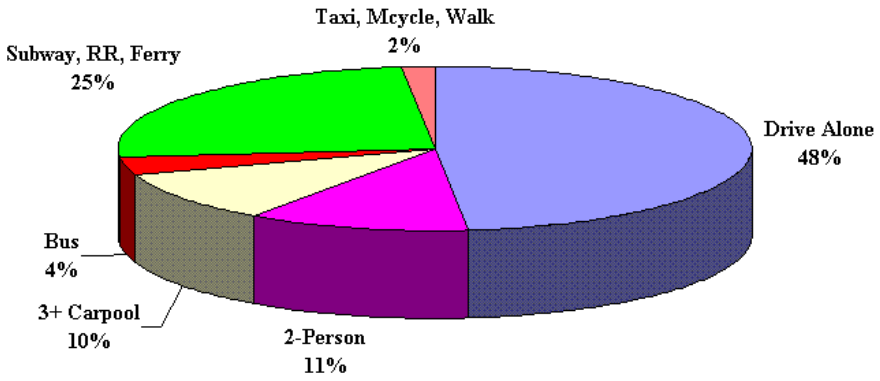

Prince William Co. CTPP Part 3 Percent Workers & Means of Transportation to Work
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
WDC
Arl
Mgm
PG
Ffx
Fqr
Ldn
PWm
Spt
Stf
FxC
FCC
FrC
MnC
MPC
Place of Work
Worker
Trans
Carpool
Alone


Aggregated # of Workers in Carpools by Time Leaving Home
Total
5-9AM
9-5AM
25000
20000
15000
10000
5000
0
Arl
Ffx
Ldn
PWm
Alx
FxC
FCC
MnC
MPC

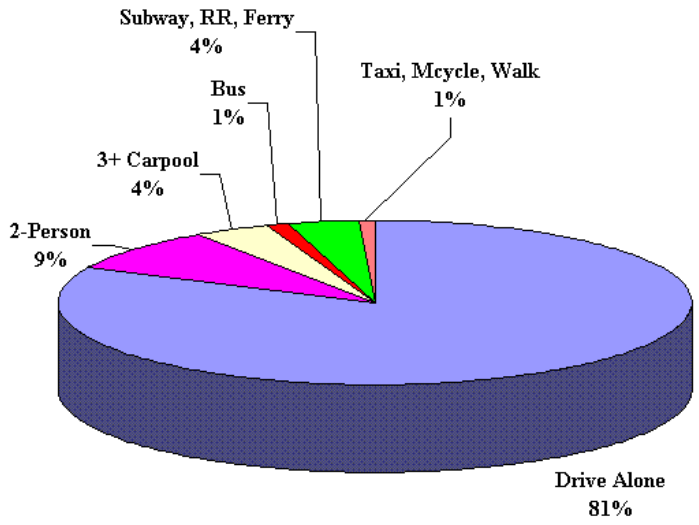

Mean # of Workers in Carpools by Time Leaving Home
Total
5-9AM
9-5AM
2.45
2.40
2.35
2.30
2.25
2.20
2.15
2.10
2.05
2.00
1.95
1.90
Arl
Ffx
Ldn
PWm
Alx
FxC
FCC
MnC
MPC

Census Transportation Planning Package, CTPP, data provides more distinctive carpool information for the undergoing *Carpooling Characteristics* study. CTTP is based on the information gathered from the decennial census and is designed for use by transportation planners. However, PUMS files are perfect for people, such as data users in academic life -- economists, psychologists, and sociologists – which have found the PUMS useful for regression analysis and modeling applications. The following data shown in pie charts are more in general form and could be available by analyzing either PUMS or CTPP data.

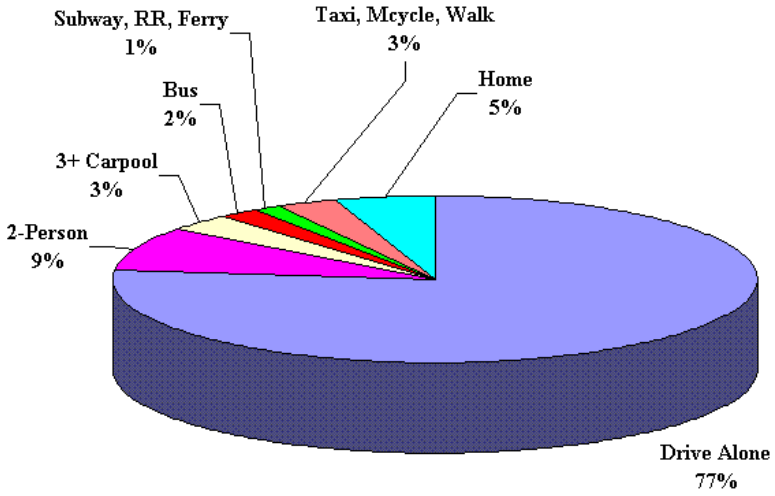
Common data production utilizing IPUMS & CTPP



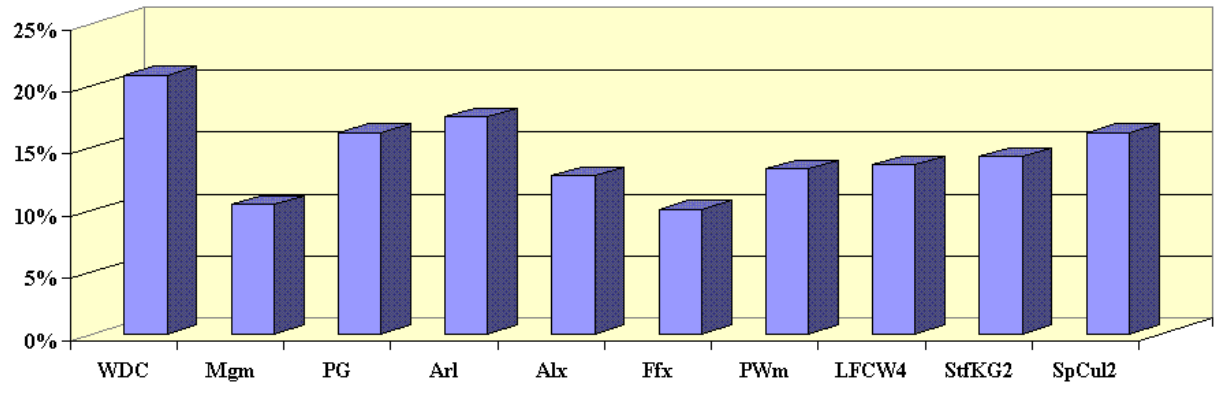
NOVA to MD Work-flow by Means of Transportation



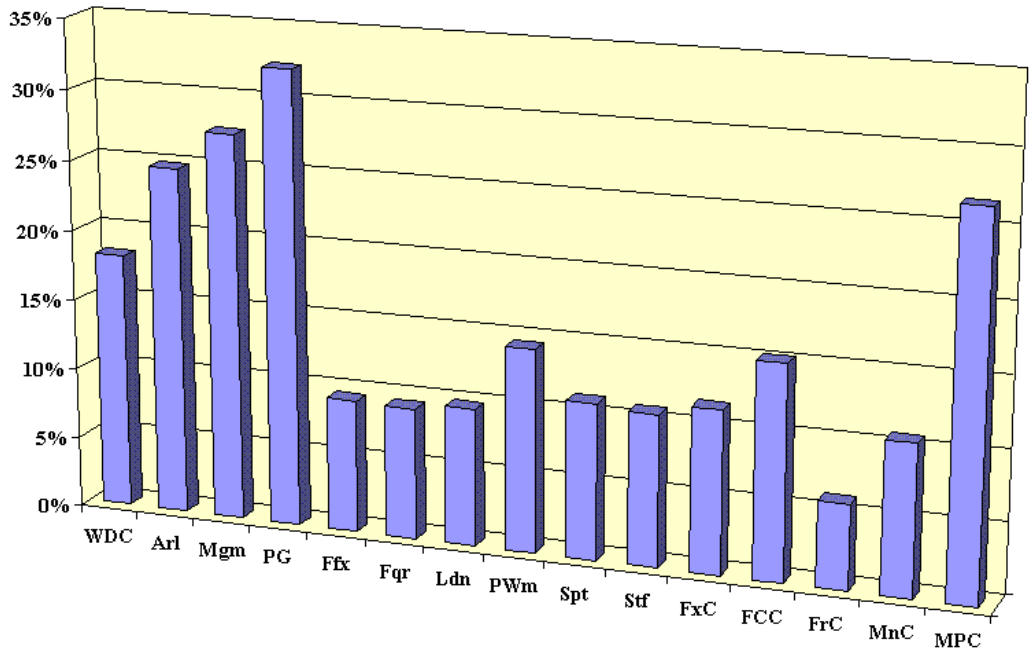
NOVA to NOVA Work-flow by Means of Transportation



PUMS vs. CTPP


NOVA PUMS Carpoolers by Place of Work
25%
20%
15%
10%
5%
0%
WDC
Mgm
PG
Arl
Alx
Ffx
PWm
LFCW4
StfKG2
SpCul2

NOVA CTPP Part-3 Carpoolers by Place of Work



Conclusion

2000 Census PUMS & IPUMS samples of data are code structure and require an extensive analysis. Best approach is to analyze it by "Frequency". Then, codes can be converted for the final product. IPUMS gives you the flexibility of tailoring the data records by Variables. For the Transportation Engineers, IPUMS is a better option for the source of data, in the absence of CTPP data. However, CTPP data are more specific in Place of Resident & Place of Work. Should I wait for the CTPP data? For the Metro/County-wide studies, I shouldn't. For more specific & localized studies, I should. For *Carpooling Characteristics* project, indeed.