Problem Set #1: Data Section

MACS 30200, Dr. Evans

Dongping Zhang

U.S. Current Population Survey (CPS) is a voluntary monthly survey collected from approximately 150,000 to U.S. households by the Bureau of Census for the Bureau of Labor Statistics. The sampling units of the survey are households and the observational units are individuals in sampled households. Census Bureau employees collect data through phone interviews by calling the number associated with the selected households and asking questions regarding demographics and employment status of each person in the selected households. A household selected for CPS would be interviewed for four consecutive months and then be dropped out of the sample for the next eight consecutive months and would be brought back in the following four months.

In addition to the national unemployment rate, CPS provides a comprehensive body of data on the labor force, individual's hours of work and earnings, together with other demographic and labor force characteristics. Using CPS, BLS analyzes the data and releases U.S. Government's official estimates of monthly unemployment rate and labor force participation rate of the country. Official CPS survey data could be accessed from three sources, and they are The Bureau of Census, The Bureau of Labor Statistics, and The National Bureau of Economic Research CPS Home Page. Among those three sources, the Bureau of Census is the original source and curator of CPS while NBER maintains an extensive collection of historical CPS public use tapes for its research staff.

The dataset described in this section is the CPS monthly survey data from January 2016, which is a fixed-width format text file with the actual data. Table 1 contains some important descriptive statistics for eight key variables, and those descriptive

Table 1: Summary statistics of characteristic and employment variables

Variable Name	PRTAGE	HRNUMHOU	PUHROFF2	PUHROT1	PEHRUSL1	PEHRUSL2	PEHRUSLT	PELAYDUR
Variables	Age	Household Size	Hours/week Takeoff	Hours/week Overtime	Hours/week on Main Job	Hours/week on Other Job	Total Hours/week in All Jobs	Duration of Layoff
min	0	0	0	1.18	0	0	1	0
max	85	16	88	72	99	81	150	119
25%	10	2	1	1	1	0	1	1
50%	33	3	1	1	1	1	1	1
75%	55	4	1	1	40	1	40	1
mean	33.92	2.91	1.32	1.18	15.43	1.23	15.63	1.49
std.dev	25.11	1.89	2.44	1.64	19.53	2.20	19.98	5.68
count	151010	151010	151010	151010	146733	150551	151010	151010

statistics are minimum, maximum, mean, quartiles, standard deviation, and number of observations. According to Table 1, it could be summarized that January 2016 CPS survey data was sampled from approximately 150,000 individuals and most of them lived in a standard household size of three. Most of those sampled individuals had one main full-time job and worked approximately 40 hours per week. The majority of sampled population did not work overtime or taking-off early on average, and the distribution of hours overtime and hours taking-off were skewed left according to quartiles of those two variables.

The histograms presented in Figure 1 indicate that there is a large number of 1s in the dataset, which I suspect to be some encoding issue caused by the Census Bureau when curating the data. However, regardless of that data anomaly, the percentages of sampled population falling in each age intervals were relatively evenly distributed before age 70. However, if conditioning the age variable on whether or not the individual was Hispanic, Jan 2016 CPS data contains significantly more Hispanic individual observations who were falling in the age interval between 5 and 50.

Figure 2 shows the distribution of hours individuals spent at their main full-time jobs. The histogram in Figure 2 only includes a subset of individuals who were older than sixteen. So the large percentage of zero working hours in the histogram could be potentially explained by those individuals (age > 16) were still receiving education and thus were probably not in the labor force. It is interesting to observe that Hispanic population in the sample tend to have a greater chance working at a full-time (40 hours/week) job than the rest of the racial groups combined. In addition, Table 3 and Table 4 are also presenting some interesting facts that Hispanic workers were also more likely to work over-time and less likely to take-off early, which could potentially be some interesting findings of labor characteristics for Hispanic workers.

Some key papers that have also used the CPS survey data are Card's "The Impact of the Mariel Boatlift on the Miami Labor Market", which was a paper studied

extensively on the effect of the Mariel Boatlift of 1980 on the Miami labor market, Day et al.'s "The Big Payoff: Educational Attainment and Synthetic Estimates of Work-Life Earnings. Special Studies. Current Population Reports", which was a paper studied the relationship between educational attainment and work-life earnings, and Burkhauser et al.'s "A Reassessment of the New Economics of the Minimum Wage Literature with Monthly Data from the Current Population Survey", which estimated the employment effects of federal minimum wage increases using monthly CPS data from 1979 through 1997.

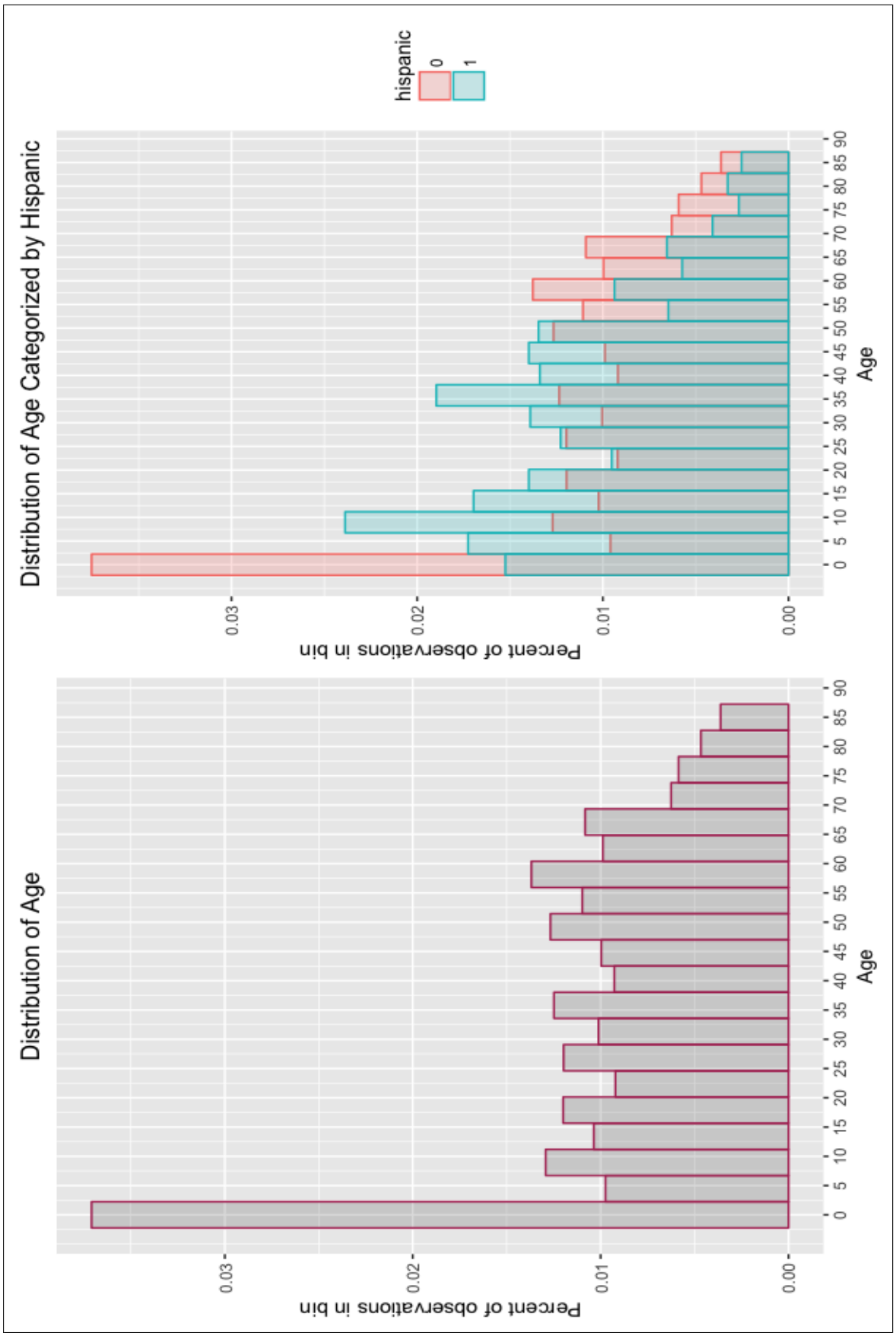

Figure 1: Distributions of Age Variable

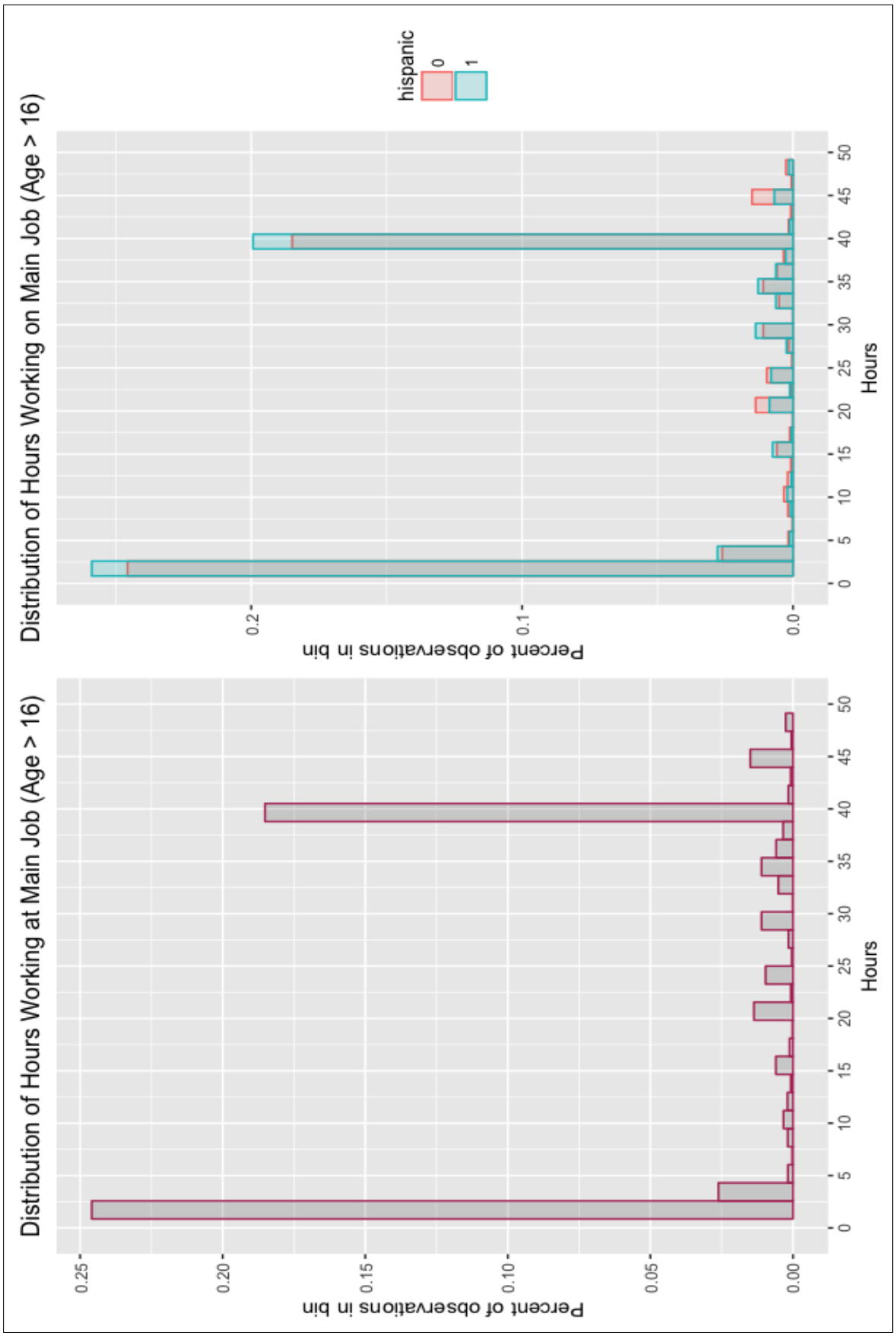

Figure 2: Distributions of Hrs. at Main Job

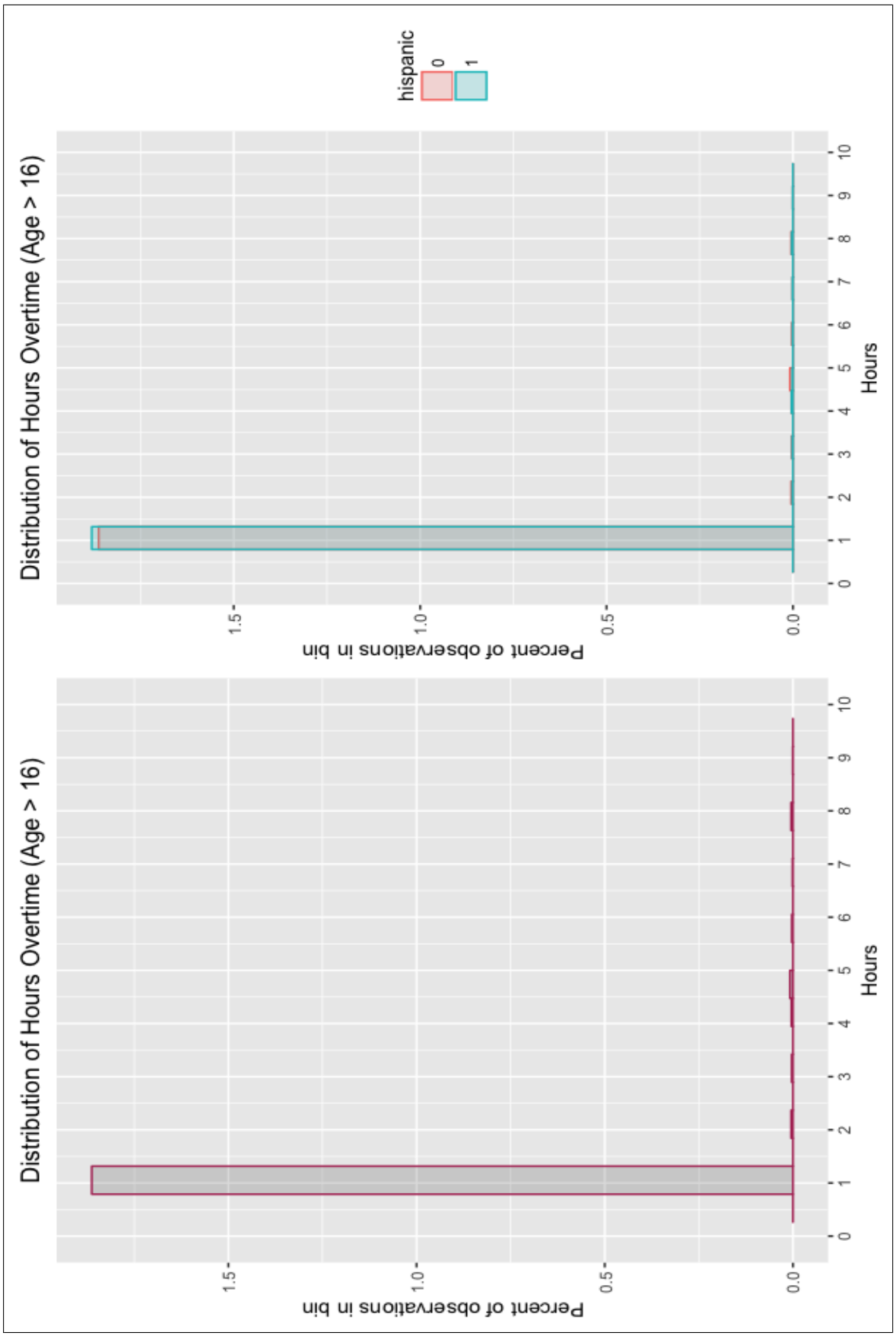

Figure 3: Distributions of Hrs. Overtime

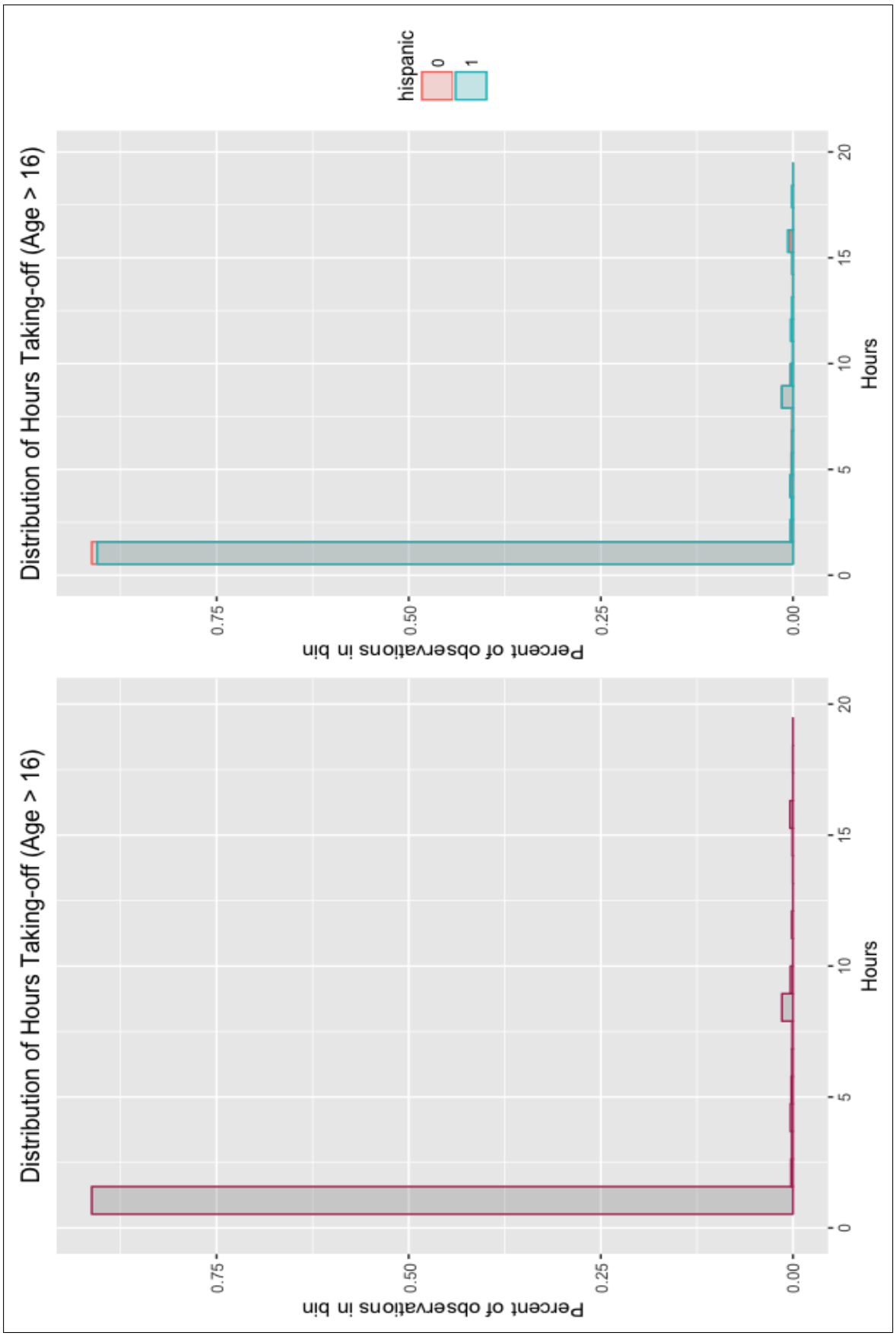

Figure 4: Distributions of Hrs. Taking-off

Reference

Burkhauser, R. V., Couch, K. A., & Wittenburg, D. C. (2000). A reassessment of the new economics of the minimum wage literature with monthly data from the Current Population Survey. *Journal of Labor Economics, 18*(4), 653-680.

Card, D. (1990). The impact of the Mariel boatlift on the Miami labor market. *ILR Review, 43*(2), 245-257.

Day, J. C., & Newburger, E. C. (2002). The Big Payoff: Educational Attainment and Synthetic Estimates of Work-Life Earnings. Special Studies. Current Population Reports.