7/16/2019 ipums.R

```
library(readr)
 1
 2
   library(dplyr)
 3 library(reshape2)
 4
   library(survey)
 5
 6
   # Export 30 from IPUMS
 7
    # "Method of travel to work, wages, occupation, industry, class of worker, and usual
    hours worked (1980, 1990, 2000-2015) w/ CSV"
 8
 9
    # Read IPUMS export
   ipums.orig <- read csv("usa 00030.csv", col types="ciiiiidi")</pre>
10
11
12
   # Filter to only full-time wageworkers (>= 35 hours per week, with wages, and not self-
    employed)
13
    ipums <- ipums.orig %>%
14
      filter(INCWAGE > 0 & INCWAGE < 999998 & CLASSWKR == 2 & UHRSWORK >= 35)
15
   # Convert years to a factor
16
17
    ipums$year <- as.factor(ipums$YEAR)</pre>
18
   # Convert TRANWORK
19
20
    ipums$commute <- NA
21
   ipums$commute[ipums$TRANWORK >= 10 & ipums$TRANWORK <= 20] <- "Private vehicle"</pre>
22
   ipums$commute[ipums$TRANWORK >= 30 & ipums$TRANWORK <= 36] <- "Public transit"</pre>
23
    ipums$commute[ipums$TRANWORK == 40] <- "Bicycle"</pre>
24
25
    ipums$commute[ipums$TRANWORK == 50] <- "Walked"</pre>
   ipums$commute[ipums$TRANWORK == 60] <- "Other"</pre>
26
    ipums$commute[ipums$TRANWORK == 70] <- "Worked at home"</pre>
27
28
    ipums$commute <- factor(ipums$commute, level=c("Private vehicle", "Public transit",</pre>
29
    "Bicycle", "Walked", "Other", "Worked at home"))
30
31
   # Recode OCC2010 to match ATUS OCC2 coding
32
   ipums$jobs <- NA
33
34
   ipums$jobs[ipums$OCC2010 >= 10 & ipums$OCC2010 <= 430] <- "Management occupations"</pre>
   ipums$jobs[ipums$OCC2010 >= 500 & ipums$OCC2010 <= 950] <- "Business and financial"</pre>
    operations occupations"
ipums$jobs[ipums$OCC2010 >= 1000 & ipums$OCC2010 <= 1240] <- "Computer and mathematical"</pre>
    science occupations"
37 | ipums$jobs[ipums$OCC2010 >= 1300 & ipums$OCC2010 <= 1560] <- "Architecture and
    engineering occupations"
38
   ipums$jobs[ipums$OCC2010 >= 1600 & ipums$OCC2010 <= 1980] <- "Life, physical, and social
    science occupations"
    ipums$jobs[ipums$OCC2010 >= 2000 & ipums$OCC2010 <= 2060] <- "Community and social
39
    service occupations"
   ipums$jobs[ipums$OCC2010 >= 2100 & ipums$OCC2010 <= 2150] <- "Legal occupations"
40
41 ipums$jobs[ipums$OCC2010 >= 2200 & ipums$OCC2010 <= 2550] <- "Education, training, and
    library occupations"
42 ipums$jobs[ipums$OCC2010 >= 2600 & ipums$OCC2010 <= 2920] <- "Arts, design,
    entertainment, sports, and media occupations"
   ipums$jobs[ipums$OCC2010 >= 3000 & ipums$OCC2010 <= 3540] <- "Healthcare practitioner and
    technical occupations"
44 ipums$jobs[ipums$OCC2010 >= 3600 & ipums$OCC2010 <= 3650] <- "Healthcare support
    occupations"
```

7/16/2019 ipums.R

```
ipums$jobs[ipums$OCC2010 >= 3700 & ipums$OCC2010 <= 3950] <- "Protective service"</pre>
45
    occupations"
    ipums$jobs[ipums$OCC2010 >= 4000 & ipums$OCC2010 <= 4150] <- "Food preparation and
    serving related occupations"
    ipums$jobs[ipums$OCC2010 >= 4200 & ipums$OCC2010 <= 4250] <- "Building and grounds"
47
    cleaning and maintenance occupations"
    ipums$jobs[ipums$OCC2010 >= 4300 & ipums$OCC2010 <= 4650] <- "Personal care and service"</pre>
48
    occupations"
    ipums$jobs[ipums$OCC2010 >= 4700 & ipums$OCC2010 <= 4965] <- "Sales and related
49
    occupations"
    ipums$jobs[ipums$OCC2010 >= 5000 & ipums$OCC2010 <= 5940] <- "Office and administrative"</pre>
50
    support occupations"
    ipums$jobs[ipums$OCC2010 >= 6005 & ipums$OCC2010 <= 6130] <- "Farming, fishing, and
    forestry occupations"
    ipums$jobs[ipums$OCC2010 >= 6200 & ipums$OCC2010 <= 6940] <- "Construction and extraction</pre>
    occupations"
    ipums$jobs[ipums$OCC2010 >= 7000 & ipums$OCC2010 <= 7630] <- "Installation, maintenance,</pre>
53
    and repair occupations"
    ipums$jobs[ipums$OCC2010 >= 7700 & ipums$OCC2010 <= 8965] <- "Production occupations"</pre>
54
    ipums$jobs[ipums$OCC2010 >= 9000 & ipums$OCC2010 <= 9750] <- "Transportation and material"</pre>
    moving occupations"
56
57
    ipums$jobs <- as.factor(ipums$jobs)</pre>
58
59
   # Read occupation codes
60
    occ <-
    read csv("https://raw.githubusercontent.com/wireservice/lookup/master/occ/description.201
    0.csv", col_types="ic")
61
    occ$description <- factor(occ$description)</pre>
62
63
64
    # Read industry codes
    ind <- read_csv("ind1990.csv", col_types="ic")</pre>
65
66
67
    ind$description <- factor(ind$description)</pre>
68
   # Read CPI rates
69
70
    cpi <-
    read csv("https://raw.githubusercontent.com/wireservice/lookup/master/year/cpi.csv",
    col types="cd")
71
72
   # Function to perform inflation adjustment on a list of columns
73
    cpi.adjust <- function(input, year.column, columns) {</pre>
74
      temp <- left join(input, cpi, by = year.column)</pre>
75
76
      for (column in columns) {
77
        temp[[column]] <- temp[[column]] * 237.0 / temp[["cpi"]]</pre>
78
      }
79
80
      temp$cpi <- NULL
81
82
      return(temp)
    }
83
84
    # Compute mean wages by year and commute
85
    # NB: Results are a perfect match for IPUMS online analysis
86
    ipums.commute.means <- ipums %>%
```

7/16/2019 ipums.R

```
88
       group by(year, commute) %>%
       summarise(
 89
 90
         pop = sum(PERWT),
 91
         mean.wages = sum(INCWAGE * PERWT) / pop,
 92
         v = (sum(PERWT * (INCWAGE - mean.wages)^2)) / (pop - 1),
 93
         sd = sqrt(v),
 94
         se = sd / sqrt(pop)
 95
       )
 96
 97
    ipums.commute.means.cpi <- cpi.adjust(ipums.means, "year", c("mean.wages", "v", "sd",</pre>
     "se"))
98
    write_csv(ipums.commute.means.cpi, "results/ipums.commute.means.cpi.csv")
99
100
101
    # Compute mean wages for home workers by year and occupation
102
    ipums.occ.means <- ipums %>%
      filter(commute == "Worked at home") %>%
103
104
       group by(year, OCC2010) %>%
105
       summarise(
106
         pop = sum(PERWT),
107
         mean.wages = sum(INCWAGE * PERWT) / pop,
         v = (sum(PERWT * (INCWAGE - mean.wages)^2)) / (pop - 1),
108
109
         sd = sqrt(v),
110
         se = sd / sqrt(pop)
111
      ) %>%
      left join(occ, by = c("OCC2010" = "occ"))
112
113
     ipums.occ.means.cpi <- cpi.adjust(ipums.occ.means, "year", c("mean.wages", "v", "sd",</pre>
114
     "se"))
115
    write_csv(ipums.occ.means.cpi, "results/ipums.occ.means.cpi.csv")
116
117
118
    # Compute mean wages for home workers by year and job category
ipums.jobs.means <- ipums %>%
120
      filter(commute == "Worked at home") %>%
121
      group by(year, jobs) %>%
122
      summarise(
123
         pop = sum(PERWT),
124
         mean.wages = sum(INCWAGE * PERWT) / pop,
125
         v = (sum(PERWT * (INCWAGE - mean.wages)^2)) / (pop - 1),
126
        sd = sqrt(v),
127
         se = sd / sqrt(pop)
128
      )
129
130
     ipums.jobs.means.cpi <- cpi.adjust(ipums.jobs.means.cpi, "year", c("mean.wages", "v",</pre>
     "sd", "se"))
131
132
    write csv(ipums.jobs.means.cpi, "results/ipums.jobs.means.cpi.csv")
133
134
    # Compute mean wages by job category and commute type for 2015 only
ipums.jobs.2015.means <- ipums %>%
136
      filter(year == 2015) %>%
137
      group_by(jobs, commute) %>%
138
      summarise(
139
         pop = sum(PERWT),
         mean.wages = sum(INCWAGE * PERWT) / pop,
140
141
         v = (sum(PERWT * (INCWAGE - mean.wages)^2)) / (pop - 1),
```

```
7/16/2019
                                                     ipums.R
 142
          sd = sqrt(v),
 143
          se = sd / sqrt(pop)
 144
 145
 146
      write csv(ipums.jobs.2015.means, "results/ipums.jobs.2015.means.csv")
 147
 148
      # Compute fraction of occupation that works from home by year
 149
      ipums.jobs.shares <- ipums %>%
 150
        group_by(year, jobs, commute) %>%
 151
        summarise(
 152
          pop = sum(PERWT)
        ) %>%
 153
 154
        mutate(
 155
          share = pop / sum(pop)
 156
        )
 157
      write csv(ipums.jobs.shares, "results/ipums.jobs.shares.csv")
 158
 159
 160
      # Compute share of homeworkers by year and industry
 161
      ipums.ind.shares <- ipums %>%
        group_by(year, IND1990, commute) %>%
 162
 163
        summarise(
          pop = sum(PERWT)
 164
 165
        ) %>%
 166
        mutate(
          share = pop / sum(pop)
 167
 168
        ) %>%
        left_join(ind, by = c("IND1990" = "ind"))
 169
 170
 171
      write_csv(ipums.ind.shares, "results/ipums.ind.shares.csv")
 172
```