Overview

- **IPP**: Division of International Prices
- **PPI**: Division of Industrial Prices and Price Indexes
- **CE**: Division of Consumer Expenditure Survey
- **OCWC**: Office of Compensation and Working Conditions
- **OSMR**: Office of Survey Methods and Research
Overview

- Automation (IPP)
- Quality control (PPI)
- Real-time response rates (OCWC)
- Data visualization (CE)
- Other R Shiny applications
- R packages
R Shiny Applications
Sample Refinement Automation

International Prices Program

- Receive data from Census and Customs
- Must verify Establishment ID Number (EIN), name, and address to provide to field economists
- 1700 export collections units per sample
- 2400 import collection units per sample
- 6 IPP sample team members
- 16 copies, 20 pastes, and 46 clicks per unit
Data Sources
Enter a Collection Unit and Sample

Collection Unit (last 4 digits):

Sample:

X43

Company Name (edited):

Corp Div:

character(0)

Street:

logical(0)

Search in New Tab

Open Google Maps

Open SOS Website

Create Future Note
### Sampled Company

<table>
<thead>
<tr>
<th>CORP_DIV</th>
<th>EIN</th>
<th>COMPANY_NAME</th>
<th>ADDRESS_1</th>
<th>CITY</th>
<th>STATE</th>
<th>ZIP_CODE_10</th>
<th>IPP_history</th>
</tr>
</thead>
</table>

**ACE:** matching on sampled EIN

<table>
<thead>
<tr>
<th>Entry.Date</th>
<th>Importer.No.</th>
<th>Importer.Name</th>
<th>Mailing.Line.1</th>
<th>M.City</th>
<th>M.St</th>
<th>M.Zip</th>
<th>Physical.Street</th>
<th>P.City</th>
<th>P.St</th>
<th>P.Zip</th>
</tr>
</thead>
</table>

**RTS Master Listing:** matching on Name, Corp Div, or Street displayed on the left
Search Results

Enter a Collection Unit and Sample
Collection Unit (last 4 digits):
0340
Sample:
M43
Company Name [edited]:
XXXXXX
Corp Div: 33142-000
Street:
G Search in New Tab
Q Open SOS Website
Create Future Note

Company Name Changes
No company name changes in the reporter or old_co_name_fn tables for Corp Div 33142-000
Google Search: company name, city, and state
Search Images Video News Shopping Maps Books

Results:
Carrabba's Italian Grill in Huntsville, AL
https://www.carrabba.com/locations/all/huntsville
Looking for a great Italian restaurant? Bring your family and friends to the Carrabba's Italian Grill location in Huntsville today and enjoy classic Italian dishes!

Ralphie May: Stand Up Live Huntsville
standuplivehuntsville.com/event?cid=476375
Ralphie May: Stand Up Live Huntsville... Stand Up Live Huntsville, Huntsville, AL... Two item minimum... to show time. Please call the Box Office at xxx xxx xxx...

Josh Blue: Stand Up Live Huntsville
standuplivehuntsville.com/event?cid=476363
Josh Blue: Stand Up Live Huntsville... Stand Up Live Huntsville, Huntsville, AL... Two item minimum... to show time. Please call the Box Office at xxx xxx xxx...

Kountry Wayne: Stand Up Live Huntsville
standuplivehuntsville.com/event?cid=480893
Kountry Wayne: Stand Up Live Huntsville... April 09, 2017 6:30 PM... Stand Up Live Huntsville, Huntsville, AL... Please call the Box Office at xxx xxx xxx...

Chingo Bling: Stand Up Live Huntsville
standuplivehuntsville.com/event?cid=481551
Chingo Bling: Stand Up Live Huntsville... Stand Up Live Huntsville, Huntsville, AL... Two item minimum... to show time. Please call the Box Office at xxx xxx xxx...

Searches related to XXXXXX YYYYYYYYYYYYYYYYY HUNTSVILLE AL 35806
carrabba's huntsville al coupons
Carrabba's Italian Grill at Huntsville
Carrabba's Italian Grill at 35801
Carrabba's Italian Grill at Huntsville
Carrabba's Italian Grill at Huntsville
Bonefish Grill Huntsville Alabama
Export Addresses at a Glance
Benefits of Automation

- 80-100 hours per sample of time savings
  - Much less clicking
  - Better and more thorough sample review
  - More time to review more problematic collection units
Sample Refinement Automation

- Ara Khatchadourian: khatchadourian.ara@bls.gov
- Rob Sutton: sutton.robert@bls.gov
Industrial Prices Visualization Dashboard
Index Comparisons
Index Review and Revision

95% Confidence Intervals of 221122 | Electric power distribution

Percent Change


Official Percent Change

EMBARGOED DATA - NOT FOR PUBLIC RELEASE
Visualization Dashboard

- Neil Wagner: wagner.neil@bls.gov
- Steve York: york.stephen@bls.gov
Interactive CE Visualization Tool
CE Public-Use Microdata (PUMD)

- Public-Use Microdata
  - Family-level characteristics
  - Expenditures by Universal Classification Code (UCC)
  - Member-level characteristics
  - Expenditures and their characteristics by type of expenditure (EXPN... > 50 files each year!)
  - And more!
Files Required for Analysis

**Family Characteristics File**
(34,177 Observations)

<table>
<thead>
<tr>
<th>nwmid</th>
<th>censusrec</th>
<th>fam_size</th>
<th>ref_titl1</th>
<th>ref_race</th>
<th>region</th>
<th>flinchcm</th>
<th>high_edu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>02792005</td>
<td>1</td>
<td>1001.000</td>
<td>Black or African American</td>
<td>South</td>
<td>Second 20 Percent</td>
<td>High School Graduate</td>
</tr>
<tr>
<td>2</td>
<td>02792005</td>
<td>1</td>
<td>1439.350</td>
<td>White</td>
<td>West</td>
<td>Third 20 Percent</td>
<td>Some College or Associate</td>
</tr>
<tr>
<td>3</td>
<td>02792005</td>
<td>1</td>
<td>10904.084</td>
<td>White</td>
<td>South</td>
<td>Second 20 Percent</td>
<td>Some College or Associate</td>
</tr>
<tr>
<td>4</td>
<td>02792005</td>
<td>1</td>
<td>21361.304</td>
<td>White</td>
<td>South</td>
<td>Third 20 Percent</td>
<td>Less than High school</td>
</tr>
<tr>
<td>5</td>
<td>02792115</td>
<td>4</td>
<td>10482.501</td>
<td>Asian</td>
<td>Northeast</td>
<td>Third 20 Percent</td>
<td>Bachelor's Degree</td>
</tr>
<tr>
<td>6</td>
<td>02792125</td>
<td>3</td>
<td>14000.429</td>
<td>White</td>
<td>Northeast</td>
<td>Lowere 20 Percent</td>
<td>Some College or Associate</td>
</tr>
<tr>
<td>7</td>
<td>02792135</td>
<td>1</td>
<td>6143.457</td>
<td>Asian</td>
<td>West</td>
<td>Fourth 20 Percent</td>
<td>Post-graduate Degree</td>
</tr>
<tr>
<td>8</td>
<td>02792225</td>
<td>1</td>
<td>13763.790</td>
<td>White</td>
<td>South</td>
<td>Lowest 20 Percent</td>
<td>Some College or Associate</td>
</tr>
<tr>
<td>9</td>
<td>02792245</td>
<td>4</td>
<td>18285.831</td>
<td>White</td>
<td>West</td>
<td>Third 20 Percent</td>
<td>Some College or Associate</td>
</tr>
<tr>
<td>10</td>
<td>02792285</td>
<td>2</td>
<td>15954.534</td>
<td>Asian</td>
<td>West</td>
<td>Second 20 Percent</td>
<td>High School Graduate</td>
</tr>
<tr>
<td>11</td>
<td>02792285</td>
<td>2</td>
<td>10761.501</td>
<td>Black or African - American</td>
<td>South</td>
<td>Third 20 Percent</td>
<td>Bachelor's Degree</td>
</tr>
<tr>
<td>12</td>
<td>02792285</td>
<td>2</td>
<td>17750.726</td>
<td>Asian</td>
<td>South</td>
<td>Fourth 20 Percent</td>
<td>Bachelor's Degree</td>
</tr>
<tr>
<td>13</td>
<td>02792285</td>
<td>5 or More</td>
<td>18626.626</td>
<td>White</td>
<td>South</td>
<td>Second 20 Percent</td>
<td>Less than High school</td>
</tr>
<tr>
<td>14</td>
<td>02792285</td>
<td>1</td>
<td>17055.513</td>
<td>Black or African - American</td>
<td>South</td>
<td>Fourth 20 Percent</td>
<td>Some College or Associate</td>
</tr>
<tr>
<td>15</td>
<td>02792285</td>
<td>2</td>
<td>18087.147</td>
<td>White</td>
<td>Northeast</td>
<td>Third 20 Percent</td>
<td>Some College or Associate</td>
</tr>
<tr>
<td>16</td>
<td>02792285</td>
<td>1</td>
<td>1171.729</td>
<td>Asian</td>
<td>West</td>
<td>Third 20 Percent</td>
<td>Bachelor's Degree</td>
</tr>
<tr>
<td>17</td>
<td>02792455</td>
<td>1</td>
<td>18004.251</td>
<td>White</td>
<td>Northeast</td>
<td>Lowest 20 Percent</td>
<td>High School Graduate</td>
</tr>
<tr>
<td>18</td>
<td>02792555</td>
<td>2</td>
<td>18876.791</td>
<td>Asian</td>
<td>West</td>
<td>Highest 20 Percent</td>
<td>Post-graduate Degree</td>
</tr>
<tr>
<td>19</td>
<td>02792555</td>
<td>5 or More</td>
<td>23000.329</td>
<td>Black or African - American</td>
<td>Midwest</td>
<td>Third 20 Percent</td>
<td>Some College or Associate</td>
</tr>
<tr>
<td>20</td>
<td>02792555</td>
<td>2</td>
<td>19092.103</td>
<td>Asian</td>
<td>West</td>
<td>Highest 20 Percent</td>
<td>Bachelor's Degree</td>
</tr>
</tbody>
</table>

**Expenditures File**
(1,720,755 Observations)

<table>
<thead>
<tr>
<th>nwmid</th>
<th>ref_moc</th>
<th>mce</th>
<th>cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>02793025</td>
<td>010110</td>
<td>550,000</td>
</tr>
<tr>
<td>2</td>
<td>02793025</td>
<td>010110</td>
<td>253,357</td>
</tr>
<tr>
<td>3</td>
<td>02793175</td>
<td>010110</td>
<td>137,000</td>
</tr>
<tr>
<td>4</td>
<td>02793215</td>
<td>010110</td>
<td>42,000</td>
</tr>
<tr>
<td>5</td>
<td>02793245</td>
<td>010110</td>
<td>318,000</td>
</tr>
<tr>
<td>6</td>
<td>02793255</td>
<td>010110</td>
<td>300,000</td>
</tr>
<tr>
<td>7</td>
<td>02793265</td>
<td>010110</td>
<td>918,333</td>
</tr>
<tr>
<td>8</td>
<td>02793295</td>
<td>010110</td>
<td>275,000</td>
</tr>
<tr>
<td>9</td>
<td>02793455</td>
<td>010110</td>
<td>850,000</td>
</tr>
<tr>
<td>10</td>
<td>02793555</td>
<td>010110</td>
<td>1,500,000</td>
</tr>
<tr>
<td>11</td>
<td>02793815</td>
<td>010110</td>
<td>880,000</td>
</tr>
<tr>
<td>12</td>
<td>02793925</td>
<td>010110</td>
<td>1,050,000</td>
</tr>
<tr>
<td>13</td>
<td>02793435</td>
<td>010110</td>
<td>650,000</td>
</tr>
<tr>
<td>14</td>
<td>02793355</td>
<td>010110</td>
<td>615,000</td>
</tr>
<tr>
<td>15</td>
<td>02793425</td>
<td>010110</td>
<td>410,000</td>
</tr>
<tr>
<td>16</td>
<td>02793525</td>
<td>010110</td>
<td>140,000</td>
</tr>
<tr>
<td>17</td>
<td>02793545</td>
<td>010110</td>
<td>475,000</td>
</tr>
<tr>
<td>18</td>
<td>02793775</td>
<td>010110</td>
<td>800,000</td>
</tr>
<tr>
<td>19</td>
<td>02793805</td>
<td>010110</td>
<td>800,000</td>
</tr>
<tr>
<td>20</td>
<td>02793815</td>
<td>010110</td>
<td>800,000</td>
</tr>
<tr>
<td>21</td>
<td>02794175</td>
<td>010110</td>
<td>195,000</td>
</tr>
</tbody>
</table>
# Required Resources / Skills

---

```r
135  # Computed Annual Mean Estimates
136  #
137  # Required Resources / Skills
138  #
139  # Merge Interview CU weights and expenditures
140  int_df <- left_join(
141    fml, expend, filter(ucc, hin% getUCCs(expenditure, stub)) %>%
142    group_by(newid) %>% summarise(cost = sum(cost)),
143    by = "newid"
144  ) %>%
145  mutate_each_(
146    funs(replace(., is.na(.), 0)),
147    vars = c("cost", psst0("wtrap", str_pad(1:44, 2, "left", 0)))
148  )
149
150  # Compute an Interview annual mean estimate
151  int_gm <- int_df %>%
152    mutate(wt_cost = cost * finlwt2!) %>%
153    summarise(gm = mean(wt_cost, na.rm = TRUE) / sum(finlwt)) %>%
154    unlist() %>% unname()
155
156  # Merge Diary CU weights and expenditures
157  diag_df <- left_join(
158    fml, expend, filter(ucc, hin% getUCCs(expenditure, stub)) %>%
159    group_by(newid) %>% summarise(cost = sum(cost)),
160    by = "newid"
161  ) %>%
162  mutate_each_(
163    funs(replace(., is.na(.), 0)),
164    vars = c("cost", psst0("wtrap", str_pad(1:44, 2, "left", 0)))
165  )
166
167  # Compute a Diary annual mean estimate
168  diag_gm <- diag_df %>%
169    mutate(wt_cost = cost * finlwt2!) %>%
170    summarise(gm = mean(wt_cost, na.rm = TRUE) / sum(popwt)) %>%
171    unlist() %>% unname()
```

---

Introduction: Comparisons of reported expenditures

The Consumer Expenditure Survey (CE) program consists of two surveys, the Interview and Diary expenditures, income, and consumer unit (families and single consumers) characteristics. The survey covers U.S. consumer units (CUs), which we also refer to as households or families.

This application is intended to provide the user an introduction to CE data through an interactive tab accompanied by available tab showing comparisons of expenditures between a selected subsample of the CE data.

Click on the CE Visualization tab above to use the application.
Interactive CE Visualization Tool

Interactive CE Visualization Tool - 2015 Data

1. **Demographic Categories**
   - Region
   - Number of people in CU
   - Home Owner / Renter
   - CU income range
   - Highest level of education in the CU
   - Race of the reference person

2. **Subcategories**
   - Region
   - CU income range
   - Lowest 20 Percent
   - Race of the reference person
     - Asian

3. **Options**
   - Independent scales
   - Download Table

4. **Number of households in your sample**
   - 28

---

Housing
Food at home
Healthcare
Transportation
Utilities, fuels, and public services
Education
Cellular phone service
Apparel and services
Entertainment
Cash contributions

---

U.S. BUREAU OF LABOR STATISTICS • bls.gov
Interactive CE Visualization Tool

Demographic Categories

- Region
- Number of people in CU
- Home Owner / Renter
- CU income range
- Highest level of education in the CU
- Race of the reference person
Interactive CE Visualization Tool

Subcategories

Region
- Midwest

CU income rage
- Lowest 20 Percent

Race of the reference person
- Asian
- White
- Black or African American
- American Indian or Alaskan Native
- Asian
- Native Hawaiian or Other Pacific Islander
- Multi-race
Interactive CE Visualization Tool

Options

- Independent scales

Download Table

Number of households in your sample:

28
Interactive CE Visualization Tool

Error Bars

Mean = $30,040.00

CV = 24.12%

Sample Size = 3

Lower Bound = $15,548.70

Upper Bound = $44,531.30

<table>
<thead>
<tr>
<th>Sample</th>
<th>Expenditure</th>
<th>Cost</th>
<th>CV (%)</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subset</td>
<td>Education</td>
<td>30040.00</td>
<td>24.12</td>
<td>15548.70</td>
<td>44531.30</td>
<td>3</td>
</tr>
<tr>
<td>National</td>
<td>Education</td>
<td>5946.81</td>
<td>0.64</td>
<td>5870.69</td>
<td>6022.93</td>
<td>4024</td>
</tr>
</tbody>
</table>
Benefits to the user

- Accessibility: The user can access the app for **free** as long as they have internet access on a device with a web browser

- Usability: The user operates only the clean, user-friendly UI to get data, results, and visualizations
Interactive CE Visualization Tool

Arcenis Rojas: rojas.arcenis@bls.gov
Real-time Response Rate Tool

- Office of Compensation and Working Conditions
- Provide real-time response rates to field offices
  - Focus on problem collection areas
  - Improved sample representativity
Real-time Response Rate Tool

Response rates by region and/or establishment size

Detailed summaries for each region
Real-time Response Rate Tool

- Brandon Kopp (OSMR): kopp.brandon@bls.gov
- Randall Powers (OSMR): powers.randall@bls.gov
- Arcenis Rojas (CE): rojas.arcenis@bls.gov
Other Shiny Applications

- Choropleth maps of unemployment data (OSMR)
- Energy Information Administration analyzer (PPI)
- Text analysis Shiny App (Survey Methods)
R Packages
R Packages

- rpms: Recursive Partitioning for Modeling Survey Data package (Survey Methods)
- growfunctions: Bayesian Non-Parametric Dependent Models for Time-Indexed Functional Data package (Survey Methods)
Fits a linear model to survey data in each node obtained by recursively partitioning the data.

Adjusts for complex sample design features used to obtain the data.

Produces design-consistent coefficients to the least squares linear model between the dependent and independent variables.
rpms

- The main function returns the resulting binary tree with the linear model fit at every end-node.
- Daniell Toth (OSMR): toth.daniell@bls.gov
growfunctions

- Bayesian Non-Parametric Dependent Models for Time-Indexed Functional Data package (Survey Methods)
- Estimates a collection of time-indexed functions under either of Gaussian process (GP) or intrinsic Gaussian Markov random field (iGMRF) prior formulations
Dirichlet process mixture allows sub-groupings of the functions to share the same covariance or precision parameters.

The GP and iGMRF formulations both support any number of additive covariance or precision terms, respectively, expressing either or both of multiple trend and seasonality.
growfunctions

Terrance Savitsky (OSMR):
savitsky.terrance@bls.gov
Challenges
Challenges

- Data confidentiality
- Need for an R server to make apps/programs public
- Can only put Shiny apps on a webpage via iFrames or setting up an account on a cloud server (i.e., Digital Ocean, R Studio)
Contact Information

Arcenis Rojas
Economist
Division of Consumer Expenditure Surveys
www.bls.gov/cex
202-691-6884
rojas.arcenis@bls.gov