Source for all variables: NORC at the University of Chicago. (1998-2012). General Social Survey

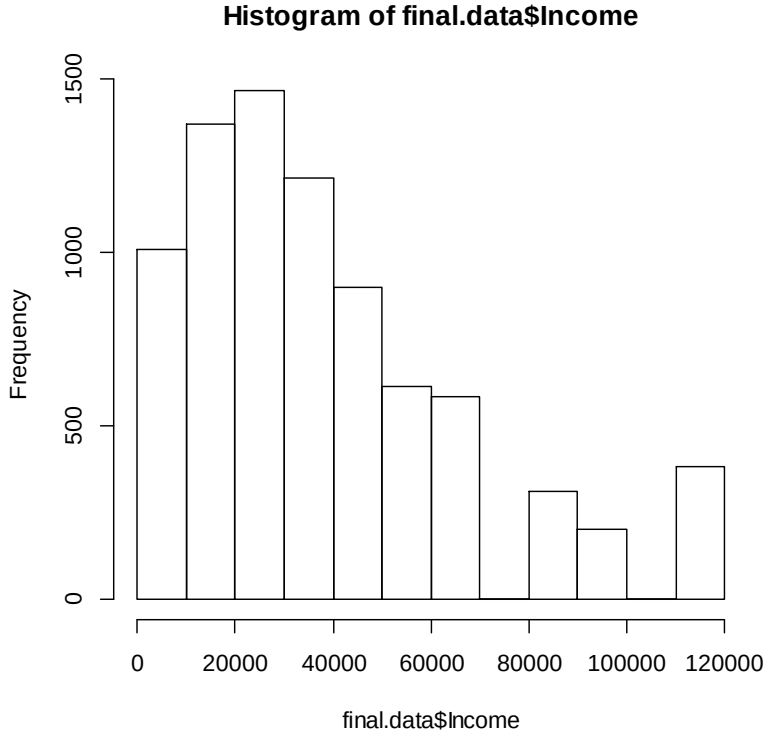
Data Appendix

Income; LogIncome:

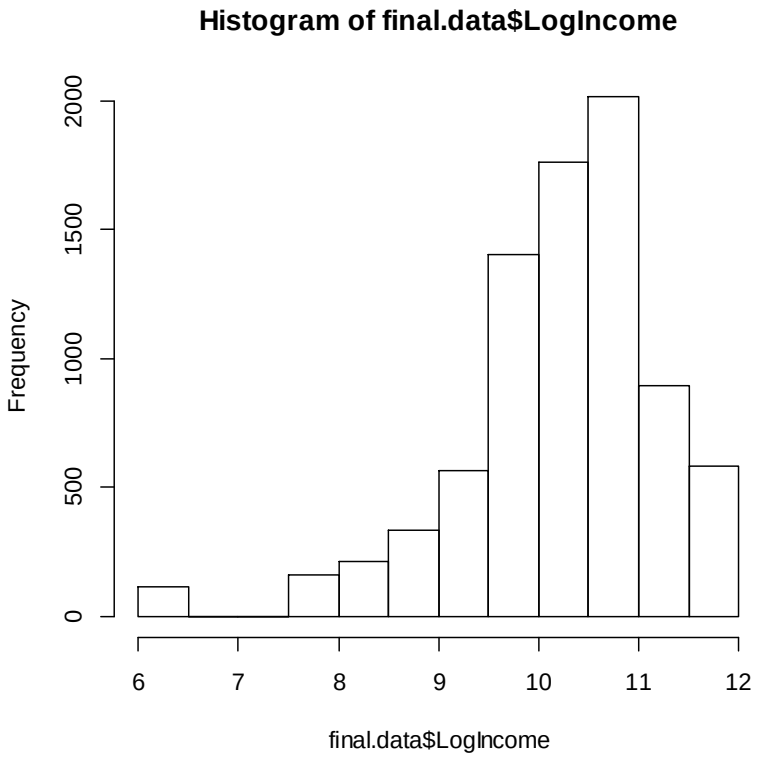
Number of observations missing: 0
Description: The variable *Income* is the individual respondent's annual income earned from his/her occupations before taxes or other deductions, measured in dollars. Annual income reported in bracketed categories, so the data uses the median amount in each income bracket. Data merged between income reported in 1998-2004 income brackets and income reported in 2006-2012 income brackets. The latter income brackets simply include the ranges: $110,000 to $129,999; $130,000 to $149,000; and $150,000 or over. Respondents from 2006-2012 in these income brackets were all placed in the final 1998-2004 income bracket, $110,000 or over. $120,000 was arbitrarily chosen to represent the "$110,000 or over" range, while $500 was arbitrarily chosen to represent the "Under $1,000" range. *LogIncome* is the natural log of income.
Name of variable used in the source data: *rincom98, rincom06*
Coding scheme: numbers 1-23 for *rincom98*; numbers 1-25 for *rincom06*

Variable	Frequency	Mean	Median	Std. Dev.	Min	Max
Annual Income	8,053	$38,330	$32,500	$29,092.30	$500	$120,000

Histogram of Respondent's Annual Income:



Histogram of the Natural Log of Respondent's Annual Income:

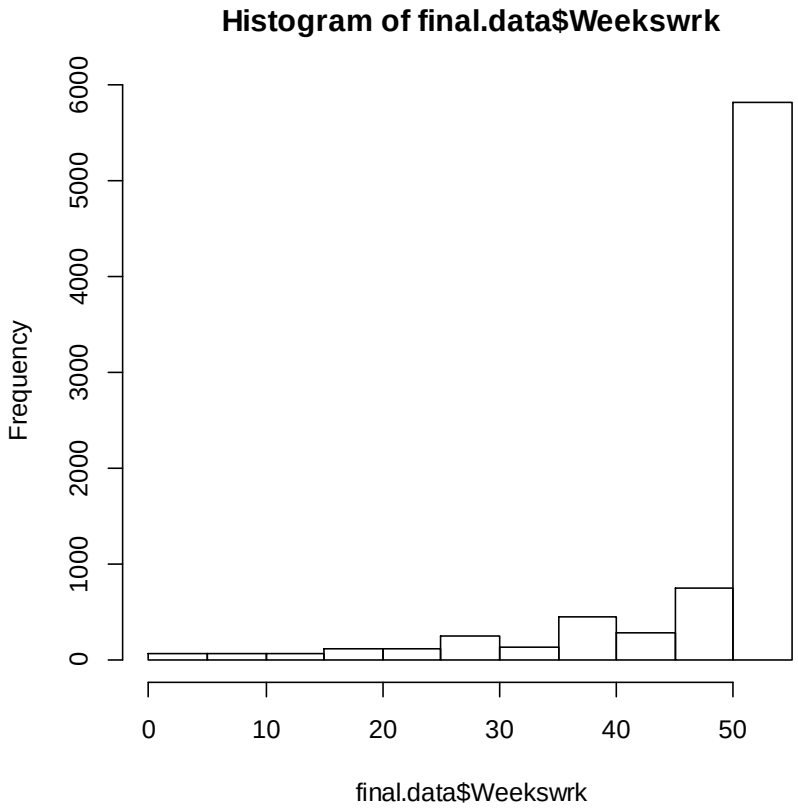


Weekswrk:

Number of observations missing: 0
Description: The number of weeks respondent worked in the previous year either full-time or part-time, including paid vacations and sick leave, measured in weeks.
Name of variable used in the source data: *weekswrk*

Variable	Frequency	Mean	Median	Std. Dev.	Min	Max
Weeks Worked	8,053	47.8	52	9.447455	0	52

Histogram of the Number of Weeks the Respondent Worked in the Previous Year:

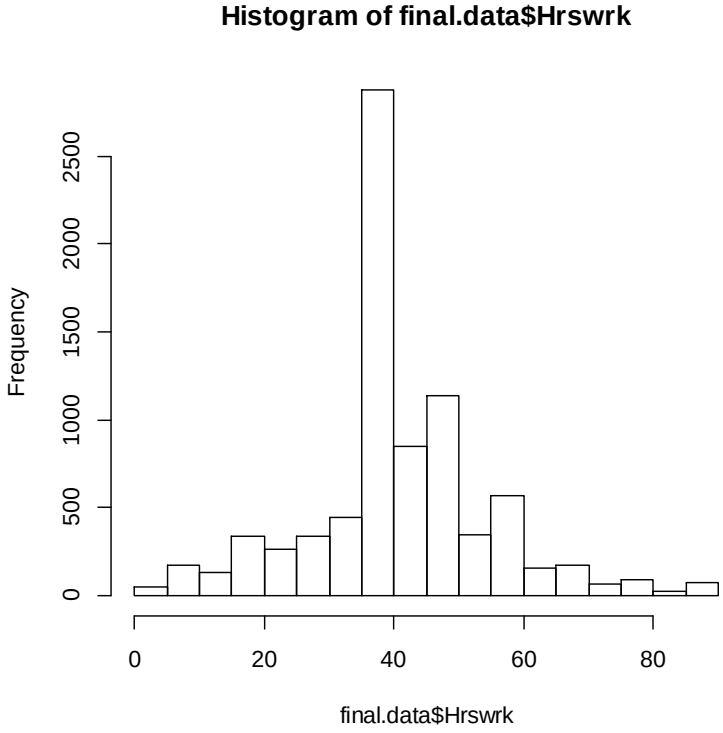


Hrswrk:

Number of observations missing: 0
Description: The number of hours the respondent worked last week, at all jobs, measured in hours. *Hourswrk* is used to proxy the number of hours the respondent usually works each week.
Name of variable used in the source data: *hrs1*

Variable	Frequency	Mean	Median	Std. Dev.	Min	Max
Hours	8,053	42.41	40	14.09647	1	89

Histogram of the Number of Hours Respondent Worked Last Week:



Female:

Number of observations missing: 0
Description: The sex of the respondent, coded as female.
Name of variable used in the source data: *sex*
Coding scheme: *female* is coded as 1 if *sex* is coded as 2, and 0 otherwise.

	Frequency	Percent Frequency
0	4,134	51.33
1	3,919	48.67
Total	8,053	100.00

Queer:

Number of observations missing: 0
Description: *Queer* is the number of respondents in the past five years who have had exclusively same-sex sex partners or both male and female sex partners. *Queer* is used to proxy for sexual orientation.
Name of variable used in the source data: *sex, sexsex5*
Coding scheme: *Queer* is coded as 1 if the respondent has had exclusively same-sex sex partners or sex partners of both sexes in the past five years, and 0 otherwise.

	Frequency	Percent Frequency
0	7,690	95.49
1	363	4.51
Total	8,053	100.00

Nonwhite:

Number of observations missing: 0
Description: The race of the respondent, separated between white and nonwhite individuals.
Name of variable used in the source data: *race*
Coding scheme: *Nonwhite* is coded as 1 if *race* is coded as 2 or 3, and 0 otherwise.

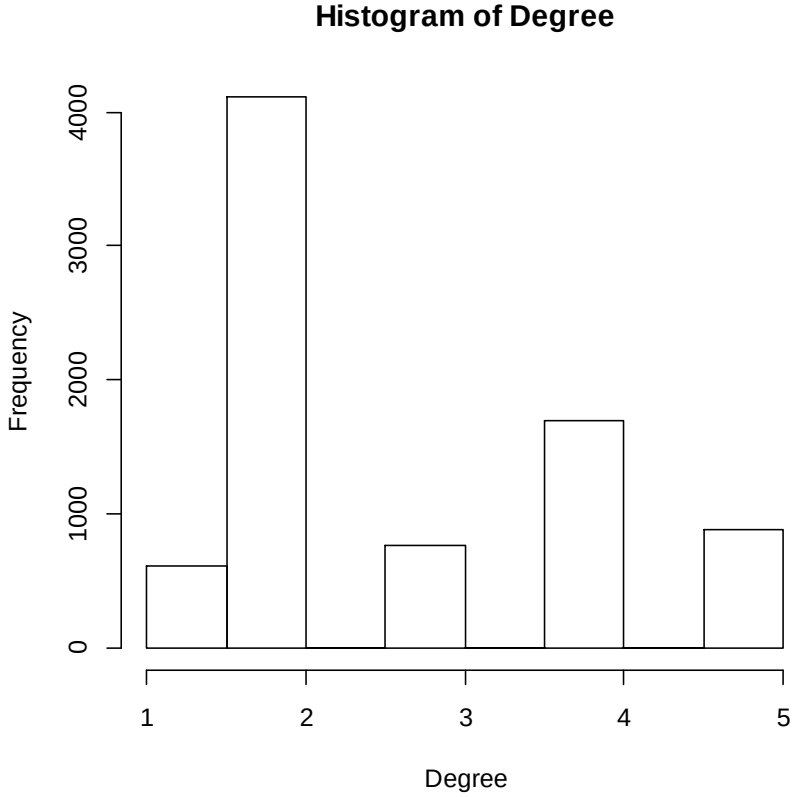
	Frequency	Percent Frequency
0	6,358	78.95
1	1,695	21.05
Total	8,053	100.00

Lesshs; Hs; Assoc; Bach; Grad:

Number of observations missing: 0
Description: Dummies representing the highest degree attained by the respondent.
Name of variable used in the source data: *degree*
Coding scheme: *Lesshs* is coded as 1 if *degree* is coded as 0; *Hs* is coded as 1 if *degree* is coded as 1; *Assoc* is coded as 1 if *degree* is coded as 2; *Bach* is coded as 1 if *degree* is coded as 3; *Grad* is coded as 1 if *degree* is coded as 4; and 0 otherwise

	Frequency	Percent Frequency
Less Than High School	610	7.57
High School	4,111	51.05
Junior College	760	9.44
Bachelor	1,692	21.01
Graduate	880	10.93
Total	8,053	100.00

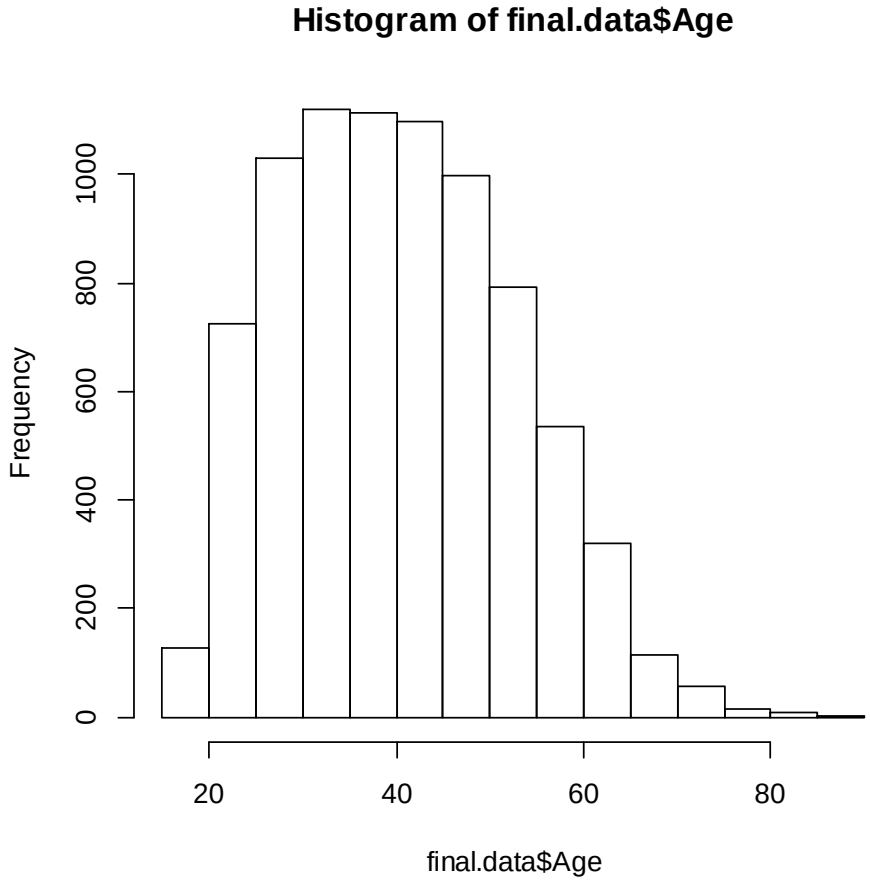
Histogram of Highest Educational Degree Attained by Respondent:



Age:

Number of observations missing: 0
Description: The age of the respondent, measured in years.
Name of variable used in the source data: *age*

Variable	Frequency	Mean	Median	Std. Dev.	Min	Max
Age	8,053	40.9	40	12.21301	18	88

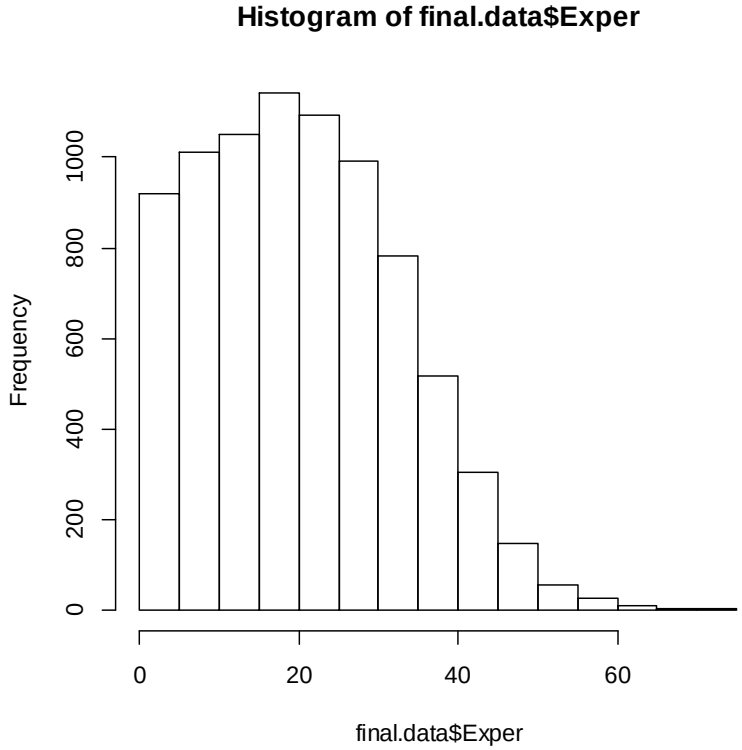


Exper; Exper_sq:

Number of observations missing: 0
Description: *Exper* is a proxy for years of work experience, measured as age minus years of education minus six. *Exper_sq* is the square of *Exper*.
Name of variable used in the source data: *age, educ*

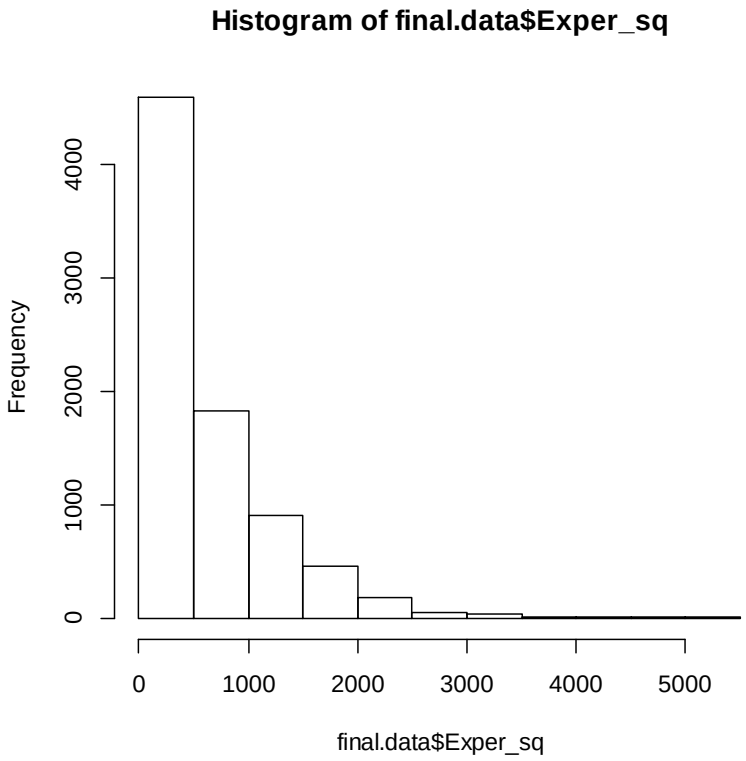
Variable	Frequency	Mean	Median	Std. Dev.	Min	Max
Potential Experience	8,053	20.87	20	12.35557	0	73

Histogram of Respondent's Potential Experience:



Variable	Frequency	Mean	Median	Std. Dev.	Min	Max
Potential Experience Squared	8,053	588	400	607.9405	0	5329

Histogram of Respondent's Potential Experience Squared:



Married:

Number of observations missing: 0
Description: Marital status of all respondents, separated between the number of individuals that are currently married and currently single (widowed, divorced, separated, or never married).
Name of variable used in the source data: *marital*
Coding scheme: *Married* is coded as 1 if the respondent is currently married, and 0 otherwise.

	Frequency	Percent Frequency
0	3794	47.11
1	4259	52.89
Total	8,053	100.00

Metro:

Number of observations missing: 0
Description: The number of respondents living in a metropolitan area (a city with at least 250,000 residents).
Name of variable used in the source data: *xnorcsiz*
Coding scheme: *Metro* is coded as 1 if the respondent is living in a metropolitan area, and 0 otherwise.

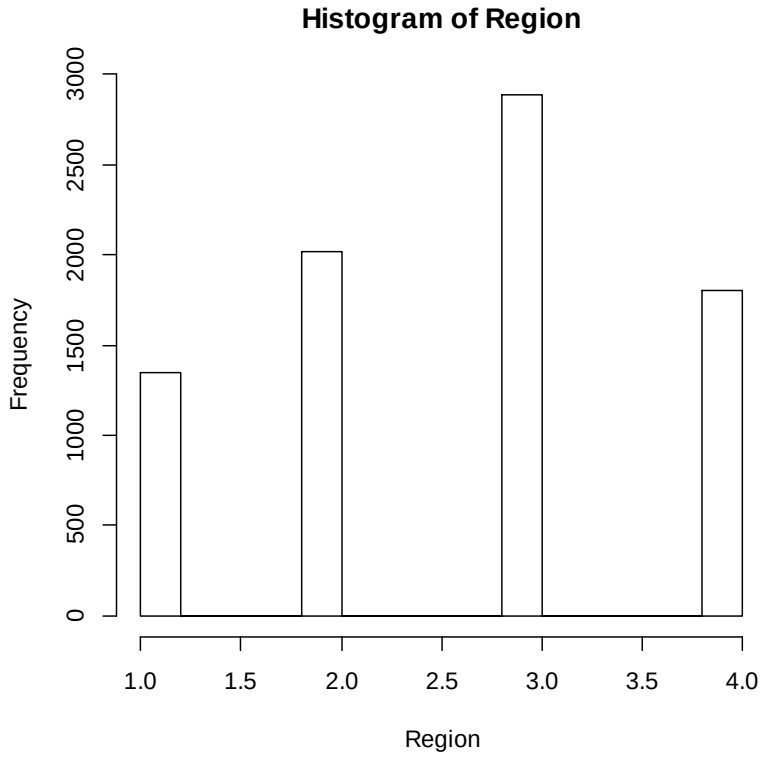
	Frequency	Percent Frequency
0	6,846	85.01
1	1,207	14.99
Total	8,053	100.00

Northeast; Midwest; South; West:

Number of observations missing: 0
Description: Dummies for the US geographic region that the respondent currently resides in.
Name of variable used in the source data: *region*
Coding scheme: *Northeast* is coded as 1 if *region* is coded as 1 or 2 (New England or Middle Atlantic); *Midwest* is coded as 1 if *region* is coded as 3 or 4 (East North Central or West North Central); *South* is coded as 1 if *region* is coded as 5, 6, or 7 (South Atlantic, East South Central, or West South Central); *West* is coded as 1 if *region* is coded as 8 or 9 (Mountain or Pacific)

	Frequency	Percent Frequency
Northeast	1,345	16.70
Midwest	2,021	25.10
South	2,886	35.84
West	1,801	22.36
Total	8,053	100.00

Histogram of US Geographic Regions of Respondents:



Person; Prof; Finance; Trade; Public; Enter; Trans; Busi; Manu; Agri; Mine; Construct:
Additional source: GSS Codebook for Cumulative Data (1972-2012). Appendix H – Industrial Classifications Distributions
Number of observations missing: 0
Description: The number of respondents in each industry, coded as dummies.
Name of variable used in the source data: *indus80, indus10*
Coding scheme: Agri is coded as 1 if indus10 is coded as 170 – 0290 or if indus80 is coded as 10 – 031; Mine is coded as 1 if indus10 is coded as 370 – 0490 or if indus80 is coded as 40 – 050; construct is coded as 1 if indus10 is coded as 770 or if indus80 is coded as 60; Manu is coded as 1 if indus10 is coded as 1070 – 3990 or 6470 – 6490 or if indus80 is coded as 100 – 392; Trans is coded as 1 if indus10 is coded as 570 – 0690, 6070 – 6390, or 6570 – 6690 or if indus80 is coded as 400 – 472; Trade is coded as 1 if indus10 is coded as 4070 – 5790 or 8680 – 8690 or if indus80 is coded as 500 – 691; Finance is coded as 1 if indus10 is coded as 6870 – 7190 or if indus80 is coded as 700 – 712; Busi is coded as 1 if indus10 is coded as 7570, 7770 – 7790, 7580 – 7780, or 8770 – 8870 or if indus80 is coded as 721 – 760; Person is coded as 1 if indus10 is coded as 8660 – 8670 or 8880 – 9090 or if indus80 is coded as 761 – 791; Enter is coded as 1 if indus10 is coded as 8560 – 8590 or if indus80 is coded as 800 – 802; Prof is coded as 1 if indus10 is coded as 6770 – 6780, 7270 – 7490, 7860 – 8470, or 9160 – 9190 or if indus80 is coded as 812 – 892; Public is coded as 1 if indus10 is coded as 9370 – 9590 or 9670 – 9870 or if indus80 is coded as 900 - 932

	Frequency	Percent Frequency
Personal Services	269	3.34
Professional And Related Services	2,229	27.68
Finance, Insurance, And Real Estate	541	6.72
Wholesale And Retail Trade	1,376	17.09
Public Administration	534	6.63
Entertainment And Recreation Services	154	1.91
Transportation, Communications, And Other Public Utilities	630	7.82
Business And Repair Services	529	6.57
Manufacturing	1,096	13.61
Agriculture	154	1.91
Mining	32	0.40
Construction	509	6.32
Total	8,053	100.00

Manprof; Service; Office; Natconst; Prodlabor:

Additional source: GSS Codebook for Cumulative Data (1972-2012). Appendix F – Occupational Classifications Distributions
Number of observations missing: 0
Description: The number of respondents in each occupational category, coded as dummies.
Name of variable used in the source data: *occ80, occ10*
Coding scheme: Manprof is coded as 1 if occ10 is coded as 10-3540 or if occ80 is coded as 3-235 or 473-476; Service is coded as 1 if occ10 is coded as 3600-4650, or if occ80 is coded as 403-469 or 487. Office is coded as 1 if occ10 is coded as 4700-5940 or if occ80 is coded as 243-389; Natconst is coded as 1 if occ10 is coded as 6000-7630 or if occ80 is coded as 477-486, 488-496, or 498-616. Prodlabor is coded as 1 if occ10 is coded as 7700-9750 or if occ80 is coded as 617-889 or 497.

	Frequency	Percent Frequency
Management, Professional, & Related Occupations	3,354	41.64
Service Occupations	1,153	14.32
Sales & Office Occupations	1.773	22.02
Natural Resources, Construction, & Maintenance Occupations	668	8.30
Production, Transportation, & Material Moving Occupations	1,105	13.72
Total	8,053	100.00