

# BLS - QCEW - Download

Datasets

*François Geerolf*

## Contents

<b>Preamble</b>	<b>1</b>
<b>Presentation of QCEW</b>	<b>1</b>
Source . . . . .	1
Layout . . . . .	2
Parsing the QCEW Webpage . . . . .	2
<b>Auxiliary files</b>	<b>6</b>
Crosswalk NAICS 2002 to SIC 1987 . . . . .	6
NAICS 2007 Codes . . . . .	6
Categories . . . . .	7
<b>Main Files</b>	<b>7</b>
NAICS Classification . . . . .	7
SIC Classification . . . . .	9
<b>Computing Environment</b>	<b>10</b>

## Preamble

```
rm(list = ls())
pklist <- c("tidyverse", "rvest", "curl", "gdata")
source("https://fgeerolf.github.io/datasets/load-packages.R")
options(tibble.print_max = 100)
```

## Presentation of QCEW

### Source

Data on employment by county and industry come from the Bureau of Labor Statistics Longitudinal Database (LDB) and **Quarterly Census of Employment and Wages (QCEW)**. The LDB reports employment by establishment and month starting in 1990.

The source data come from **quarterly reports** employers file with **state employment security** agencies as part of the unemployment insurance system; as a result, the LDB contains essentially universal coverage of private sector employment. Each establishment in the LDB has a 6 digit NAICS code associated with its primary activity.

The QCEW is the public use version of the LDB. It reports monthly employment at the industry-county level for all 50 states starting in 1975, subject to disclosure limitations to prevent the release of identifying information regarding single establishments.

Even at the NAICS 2 digit level and with counties already aggregated into metropolitan statistical areas (MSAs), roughly one-fifth of potential cells get suppressed for disclosure reasons; the suppressed share rises to 35% for MSA-industry cells at the NAICS 3 digit level.

## Layout

Flat files available here: <https://www.bls.gov/cew/datatoc.htm>.

NAICS-based file layout is here: [https://data.bls.gov/cew/doc/layouts/csv\\_quarterly\\_layout.htm](https://data.bls.gov/cew/doc/layouts/csv_quarterly_layout.htm) SIC-based file layout is here: [https://data.bls.gov/cew/doc/layouts/sic\\_csv\\_quarterly\\_layout.htm](https://data.bls.gov/cew/doc/layouts/sic_csv_quarterly_layout.htm)

- SIC Layout

```
sic.layout <- read.csv("https://data.bls.gov/cew/doc/layouts/sic_csv_quarterly_layout.csv")
save(sic.layout, file = "sic.layout.RData")
```

- NAICS Layout

```
naics.layout <- read.csv("https://data.bls.gov/cew/doc/layouts/csv_quarterly_layout.csv")
save(naics.layout, file = "naics.layout.RData")
```

## Parsing the QCEW Webpage

The NAICS Table is the first table on the webpage:

```
xml2::read_html("https://www.bls.gov/cew/datatoc.htm") %>%
  html_table(., header = TRUE)
```

```
## [[1]]
##      Excel Files CSVsBy Area      CSVsBy Area CSVsBy Industry
## 1 CountyHigh-Level   Quarterly AnnualAverages      Quarterly
## 2      File Layout File Layout      File Layout      File Layout
## 3          2018      2018          N/A          2018
## 4          2017      2017          2017          2017
## 5          2016      2016          2016          2016
## 6          2015      2015          2015          2015
## 7          2014      2014          2014          2014
## 8          2013      2013          2013          2013
## 9          2012      2012          2012          2012
## 10         2011      2011          2011          2011
## 11         2010      2010          2010          2010
## 12         2009      2009          2009          2009
## 13         2008      2008          2008          2008
## 14         2007      2007          2007          2007
## 15         2006      2006          2006          2006
## 16         2005      2005          2005          2005
## 17         2004      2004          2004          2004
## 18         2003      2003          2003          2003
## 19         2002      2002          2002          2002
## 20         2001      2001          2001          2001
## 21         2000      2000          2000          2000
## 22         1999      1999          1999          1999
## 23         1998      1998          1998          1998
## 24         1997      1997          1997          1997
## 25         1996      1996          1996          1996
```

## 26	1995	1995	1995	1995
## 27	1994	1994	1994	1994
## 28	1993	1993	1993	1993
## 29	1992	1992	1992	1992
## 30	1991	1991	1991	1991
## 31	1990	1990	1990	1990
## 32				1989
## 33				1988
## 34				1987
## 35				1986
## 36				1985
## 37				1984
## 38				1983
## 39				1982
## 40				1981
## 41				1980
## 42				1979
## 43				1978
## 44				1977
## 45				1976
## 46				1975
##	CSVsBy Industry	CSVsSingle Files	CSVsSingle Files	CSVsBy Size
## 1	AnnualAverages	Quarterly	AnnualAverages	FirstQuarter
## 2	File Layout	File Layout	File Layout	File Layout
## 3	N/A	2018	N/A	2018
## 4	2017	2017	2017	2017
## 5	2016	2016	2016	2016
## 6	2015	2015	2015	2015
## 7	2014	2014	2014	2014
## 8	2013	2013	2013	2013
## 9	2012	2012	2012	2012
## 10	2011	2011	2011	2011
## 11	2010	2010	2010	2010
## 12	2009	2009	2009	2009
## 13	2008	2008	2008	2008
## 14	2007	2007	2007	2007
## 15	2006	2006	2006	2006
## 16	2005	2005	2005	2005
## 17	2004	2004	2004	2004
## 18	2003	2003	2003	2003
## 19	2002	2002	2002	2002
## 20	2001	2001	2001	2001
## 21	2000	2000	2000	2000
## 22	1999	1999	1999	1999
## 23	1998	1998	1998	1998
## 24	1997	1997	1997	1997
## 25	1996	1996	1996	1996
## 26	1995	1995	1995	1995
## 27	1994	1994	1994	1994
## 28	1993	1993	1993	1993
## 29	1992	1992	1992	1992
## 30	1991	1991	1991	1991
## 31	1990	1990	1990	1990
## 32	1989			

```
## 33      1988
## 34      1987
## 35      1986
## 36      1985
## 37      1984
## 38      1983
## 39      1982
## 40      1981
## 41      1980
## 42      1979
## 43      1978
## 44      1977
## 45      1976
## 46      1975
##      LegacyFlat Files
## 1      All ENB/END
## 2      File Layouts
## 3      2018
## 4      2017
## 5      2016
## 6      2015
## 7      2014
## 8      2013
## 9      2012
## 10     2011
## 11     2010
## 12     2009
## 13     2008
## 14     2007
## 15     2006
## 16     2005
## 17     2004
## 18     2003
## 19     2002
## 20     2001
## 21     2000
## 22     1999
## 23     1998
## 24     1997
## 25     1996
## 26     1995
## 27     1994
## 28     1993
## 29     1992
## 30     1991
## 31     1990
## 32
## 33
## 34
## 35
## 36
## 37
## 38
## 39
```

```

## 40
## 41
## 42
## 43
## 44
## 45
## 46
##
## [[2]]
##      CSVsBy Area      CSVsBy Area CSVsBy Industry CSVsBy Industry
## 1      Quarterly AnnualAverages      Quarterly      AnnualAverages
## 2      File Layout      File Layout      File Layout      File Layout
## 3          2000          2000          2000          2000
## 4          1999          1999          1999          1999
## 5          1998          1998          1998          1998
## 6          1997          1997          1997          1997
## 7          1996          1996          1996          1996
## 8          1995          1995          1995          1995
## 9          1994          1994          1994          1994
## 10         1993          1993          1993          1993
## 11         1992          1992          1992          1992
## 12         1991          1991          1991          1991
## 13         1990          1990          1990          1990
## 14         1989          1989          1989          1989
## 15         1988          1988          1988          1988
## 16         1987          1987          1987          1987
## 17         1986          1986          1986          1986
## 18         1985          1985          1985          1985
## 19         1984          1984          1984          1984
## 20         1983          1983          1983          1983
## 21         1982          1982          1982          1982
## 22         1981          1981          1981          1981
## 23         1980          1980          1980          1980
## 24         1979          1979          1979          1979
## 25         1978          1978          1978          1978
## 26         1977          1977          1977          1977
## 27         1976          1976          1976          1976
## 28         1975          1975          1975          1975
##      CSVsSingle Files CSVsSingle Files  CSVs By Size Legacy  Flat Files
## 1      Quarterly      AnnualAverages First Quarter      All EWB
## 2      File Layout      File Layout      File Layout      File Layout
## 3          2000          2000          2000          2000
## 4          1999          1999          1999          1999
## 5          1998          1998          1998          1998
## 6          1997          1997          1997          1997
## 7          1996          1996
## 8          1995          1995
## 9          1994          1994
## 10         1993          1993
## 11         1992          1992
## 12         1991
## 13         1990
## 14         1989
## 15         1988

```

## 16	1987	1987
## 17	1986	1986
## 18	1985	1985
## 19	1984	1984
## 20	1983	1983
## 21	1982	1982
## 22	1981	1981
## 23	1980	1980
## 24	1979	1979
## 25	1978	1978
## 26	1977	1977
## 27	1976	1976
## 28	1975	1975

## Auxiliary files

### Crosswalk NAICS 2002 to SIC 1987

```
curl_download(url = "https://www.census.gov/eos/www/naics/concordances/2002_NAICS_to_1987_SIC.xls",
              destfile = "2002_NAICS_to_1987_SIC.xls",
              quiet = FALSE,
              mode = "wb")

naics.sic.crosswalk <- read.xls("2002_NAICS_to_1987_SIC.xls") %>%
  select(naics = X2002.NAICS, naics.title = X2002.NAICS.Title,
         sic = SIC, sic.title = SIC.Title..and.note.) %>%
  filter(!is.na(naics)) %>%
  mutate_all(paste)

save(naics.sic.crosswalk, file = "naics.sic.crosswalk.RData")
unlink("2002_NAICS_to_1987_SIC.xls")
```

### NAICS 2007 Codes

```
curl_download(url = "https://www.census.gov/eos/www/naics/reference_files_tools/2007/naics07.xls",
              destfile = "naics07.xls",
              quiet = FALSE,
              mode = "wb")

naics.codes <- read.xls("naics07.xls") %>%
  select(-Seq..No.) %>%
  select(naics = X2007.NAICS.US.Code, naics.title = X2007.NAICS.US.Title) %>%
  filter(!is.na(naics)) %>%
  mutate_all(paste)

save(naics.codes, file = "naics.codes.RData")
unlink("naics07.xls")
```

## Categories

```
aggregation <- read.csv("https://data.bls.gov/cew/doc/titles/agglevel/agglevel_titles.csv")
save(aggregation, file = "aggregation.RData")

industry <- read.csv("https://data.bls.gov/cew/doc/titles/industry/industry_titles.csv")
save(industry, file = "industry.RData")

fips <- read.csv("https://data.bls.gov/cew/doc/titles/area/area_titles.csv")
save(fips, file = "fips.RData")

ownership <- read.csv("https://data.bls.gov/cew/doc/titles/ownership/ownership_titles.csv")
save(ownership, file = "ownership.RData")

size <- read.csv("https://data.bls.gov/cew/doc/titles/size/size_titles.csv")
save(size, file = "size.RData")
```

## Main Files

There are two periods, based on SIC or NAICS classifications:

- 1975-2000 is the **SIC** classification of industries.
- 1990-2018 is the **NAICS** classification of industries.

## NAICS Classification

1990-2017

```
for (year in c(1990:2017)){
  # year <- 2018
  cat("\nCurrently downloading BLS, NAICS", year, "\n")
  curl_download(url = paste0("https://data.bls.gov/cew/data/files/", year, "/csv/", year, "_qtrly_singlefile.csv"),
                destfile = paste0(year, "_qtrly_singlefile.zip"),
                quiet = FALSE,
                mode = "wb")

  unzip(paste0(year, "_qtrly_singlefile.zip"))
  unlink(paste0(year, "_qtrly_singlefile.zip"))

  assign(paste0("naics.", year), read.csv(paste0(year, ".q1-q4.singlefile.csv")))
  unlink(paste0(year, ".q1-q4.singlefile.csv"))

  save(list = paste0("naics.", year), file = paste0("naics.", year, ".RData"))
  rm(list = paste0("naics.", year))
}

##
## Currently downloading BLS, NAICS 1990
##
## Currently downloading BLS, NAICS 1991
##
```

```
## Currently downloading BLS, NAICS 1992
##
## Currently downloading BLS, NAICS 1993
##
## Currently downloading BLS, NAICS 1994
##
## Currently downloading BLS, NAICS 1995
##
## Currently downloading BLS, NAICS 1996
##
## Currently downloading BLS, NAICS 1997
##
## Currently downloading BLS, NAICS 1998
##
## Currently downloading BLS, NAICS 1999
##
## Currently downloading BLS, NAICS 2000
##
## Currently downloading BLS, NAICS 2001
##
## Currently downloading BLS, NAICS 2002
##
## Currently downloading BLS, NAICS 2003
##
## Currently downloading BLS, NAICS 2004
##
## Currently downloading BLS, NAICS 2005
##
## Currently downloading BLS, NAICS 2006
##
## Currently downloading BLS, NAICS 2007
##
## Currently downloading BLS, NAICS 2008
##
## Currently downloading BLS, NAICS 2009
##
## Currently downloading BLS, NAICS 2010
##
## Currently downloading BLS, NAICS 2011
##
## Currently downloading BLS, NAICS 2012
##
## Currently downloading BLS, NAICS 2013
##
## Currently downloading BLS, NAICS 2014
##
## Currently downloading BLS, NAICS 2015
##
## Currently downloading BLS, NAICS 2016
##
## Currently downloading BLS, NAICS 2017
2018:
```



```

year <- 2018
cat("\nCurrently downloading BLS, NAICS", year, "\n")

##
## Currently downloading BLS, NAICS 2018
curl_download(url = paste0("https://data.bls.gov/cew/data/files/", year, "/csv/", year, "_qtrly_singlefile.zip"),
              destfile = paste0(year, "_qtrly_singlefile.zip"),
              quiet = FALSE,
              mode = "wb")

unzip(paste0(year, "_qtrly_singlefile.zip"))
unlink(paste0(year, "_qtrly_singlefile.zip"))

assign(paste0("naics.", year), read.csv(paste0(year, ".q1-q1.singlefile.csv")))
unlink(paste0(year, ".q1-q1.singlefile.csv"))

save(list = paste0("naics.", year), file = paste0("naics.", year, ".RData"))
rm(list = paste0("naics.", year))

```

## SIC Classification

```

for (year in 1975:2000){
  # year <- 1975
  cat("\nCurrently downloading BLS, SIC", year, "\n")
  file.zip <- paste0("sic_", year, "_qtrly_singlefile.zip")
  curl_download(url = paste0("https://data.bls.gov/cew/data/files/", year, "/sic/csv/", file.zip),
                destfile = file.zip,
                quiet = FALSE,
                mode = "wb")
  unzip(file.zip)
  file.csv <- paste0("sic.", year, ".q1-q4.singlefile.csv")
  data.name <- paste0("sic.", year)
  assign(data.name, read.csv(file.csv))
  save(list = data.name, file = paste0(data.name, ".RData"))
  unlink(file.csv)
  unlink(file.zip)
  rm(data.name)
}

##
## Currently downloading BLS, SIC 1975
##
## Currently downloading BLS, SIC 1976
##
## Currently downloading BLS, SIC 1977
##
## Currently downloading BLS, SIC 1978
##
## Currently downloading BLS, SIC 1979
##
## Currently downloading BLS, SIC 1980
##

```

```
## Currently downloading BLS, SIC 1981
##
## Currently downloading BLS, SIC 1982
##
## Currently downloading BLS, SIC 1983
##
## Currently downloading BLS, SIC 1984
##
## Currently downloading BLS, SIC 1985
##
## Currently downloading BLS, SIC 1986
##
## Currently downloading BLS, SIC 1987
##
## Currently downloading BLS, SIC 1988
##
## Currently downloading BLS, SIC 1989
##
## Currently downloading BLS, SIC 1990
##
## Currently downloading BLS, SIC 1991
##
## Currently downloading BLS, SIC 1992
##
## Currently downloading BLS, SIC 1993
##
## Currently downloading BLS, SIC 1994
##
## Currently downloading BLS, SIC 1995
##
## Currently downloading BLS, SIC 1996
##
## Currently downloading BLS, SIC 1997
##
## Currently downloading BLS, SIC 1998
##
## Currently downloading BLS, SIC 1999
##
## Currently downloading BLS, SIC 2000
```

## Computing Environment

```
Sys.time()
```

```
## [1] "2018-09-26 02:23:31 PDT"
```

```
sessionInfo()
```

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.6
##
## Matrix products: default
```

```
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] bindrcpp_0.2.2  gdata_2.18.0   curl_3.2      rvest_0.3.2
## [5] xml2_1.2.0      forcats_0.3.0  stringr_1.3.1 dplyr_0.7.6
## [9] purrr_0.2.5     readr_1.1.1    tidyr_0.8.1   tibble_1.4.2
## [13] ggplot2_3.0.0   tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.18    cellranger_1.1.0 pillar_1.3.0    compiler_3.5.1
## [5] plyr_1.8.4      bindr_0.1.1     tools_3.5.1     digest_0.6.15
## [9] lubridate_1.7.4 jsonlite_1.5     evaluate_0.11   nlme_3.1-137
## [13] gtable_0.2.0    lattice_0.20-35 pkgconfig_2.0.2 rlang_0.2.2
## [17] cli_1.0.0        rstudioapi_0.7  yaml_2.2.0      haven_1.1.2
## [21] withr_2.1.2     httr_1.3.1      knitr_1.20      gtools_3.8.1
## [25] hms_0.4.2        rprojroot_1.3-2 grid_3.5.1      tidyselect_0.2.4
## [29] glue_1.3.0       R6_2.2.2        readxl_1.1.0    rmarkdown_1.10
## [33] selectr_0.4-1    modelr_0.1.2    magrittr_1.5     backports_1.1.2
## [37] scales_1.0.0     htmltools_0.3.6 assertthat_0.2.0 colorspace_1.3-2
## [41] stringi_1.2.4    lazyeval_0.2.1  munsell_0.5.0   broom_0.5.0
## [45] crayon_1.3.4
```