Snapshot S2014 availability

Lars Vilhuber

There are as of this date two versions of the MOU that binds the Census Bureau and state agencies in effect: The original MOU, signed when the state first joined the LED Program, and a successor MOU, mostly standardized across states, that states signed when the first MOU expired after 10 years. The original MOU was used for the S2008 snapshot, and subsequent snapshots have been faced with a mixture of both.

```
# version of MOU we are interested in
mouversion <- 2
mouoption    <- "A"
# this could also be "http://download.vrdc.cornell.edu/qwipu/"
urlbase <- "http://lehd.ces.census.gov/pub/"
# this could also be "latest_release"
# qwivintage <- "R2015Q3"
qwivintage <- "latest_release"
qwistates <- "AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD ME
MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI
WV WY"
qwistates <- unlist(strsplit(qwistates," "))
# common quarter to look at
# this could be deduced from metadata, here we hard-code it
qwiyear    <- 2014
qwiquarter <- 1
```

We obtained a list of the state of the MOUs as of 12-10-2015. From it, we will use the later versions (version 2) that have selected “Option A”, as those are the ones that have delegated granting access to the Census Bureau’s review process. The other options are not an ex-ante “No”, but currently, no data is available to assess whether a research project will be approved by the individual states.

We first define (source) a function to download QWI CSV files.

```
source("download_qwi.R",echo = TRUE)
```

```
##
## > download_qwi <- function(state) {
## +     qwifile <- paste("qwi", tolower(state), "sa_f_gs_ns_oslp_u",
## +         sep = "_")
## +     con <- gzcon(url(pa .... [TRUNCATED]
```

We then cycle through all the states and download the relevant file.

```
time.qwi <- system.time(for (x in qwistates) {
  eval(parse(text=paste("qwi_",tolower(x)," <- download_qwi(\"",x,"\")",sep =
""")))
  })
```

The above code can take a while. The above code ran for 19 minutes.

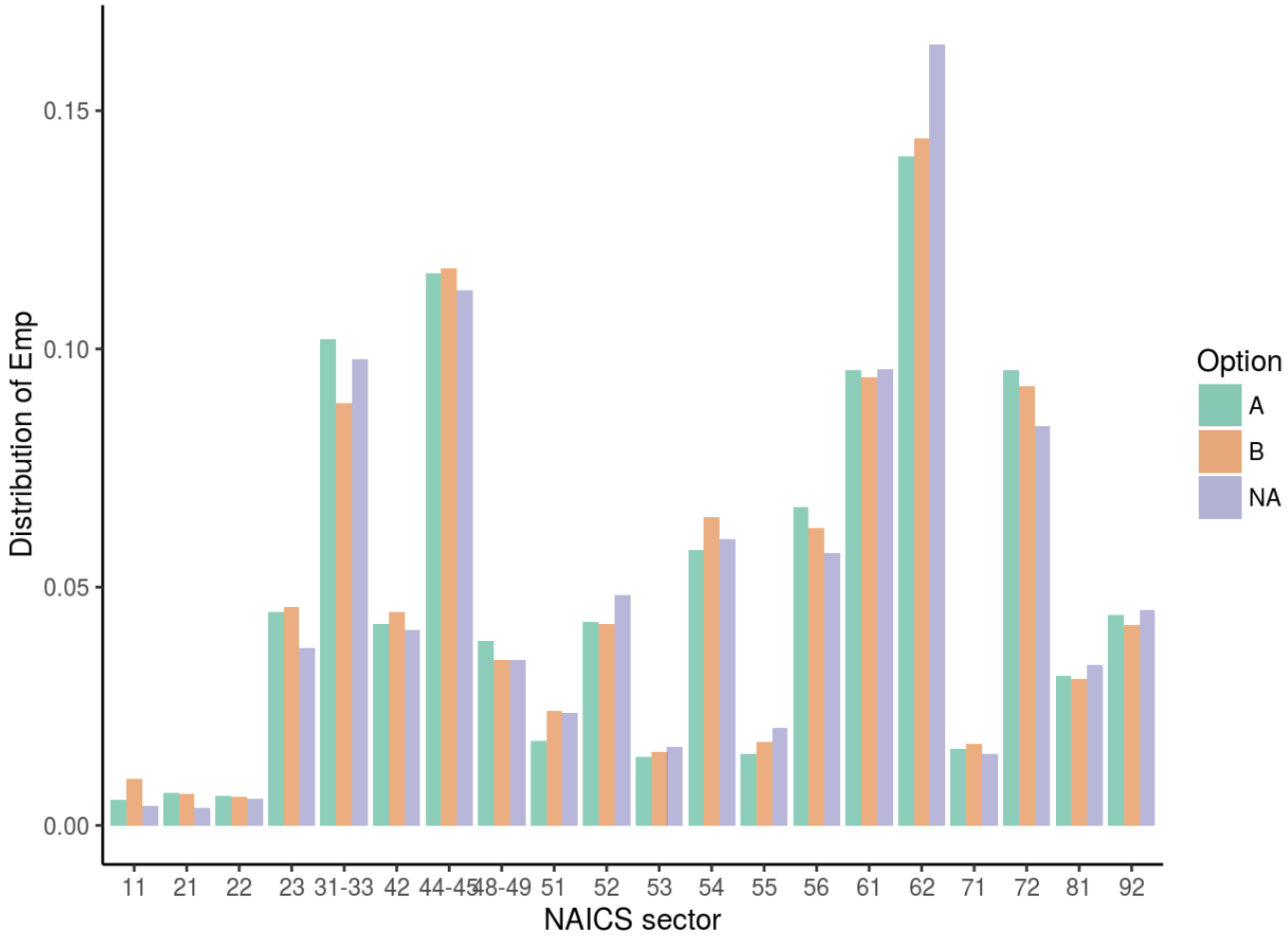
Now that we have the files, we collate them all into a single file:

```
for (x in qwistates) { eval(parse(text=paste("qwi_",tolower(x),"$state = \"",
x,"\"",sep = "")))}
for (x in qwistates[1]) { eval(parse(text=paste("all <- qwi_",tolower(x),sep =
"")))}
for (x in qwistates[-1]) { eval(parse(text=paste("all <- rbind(all,qwi_",tolower
(x),")",sep = "")))}
```

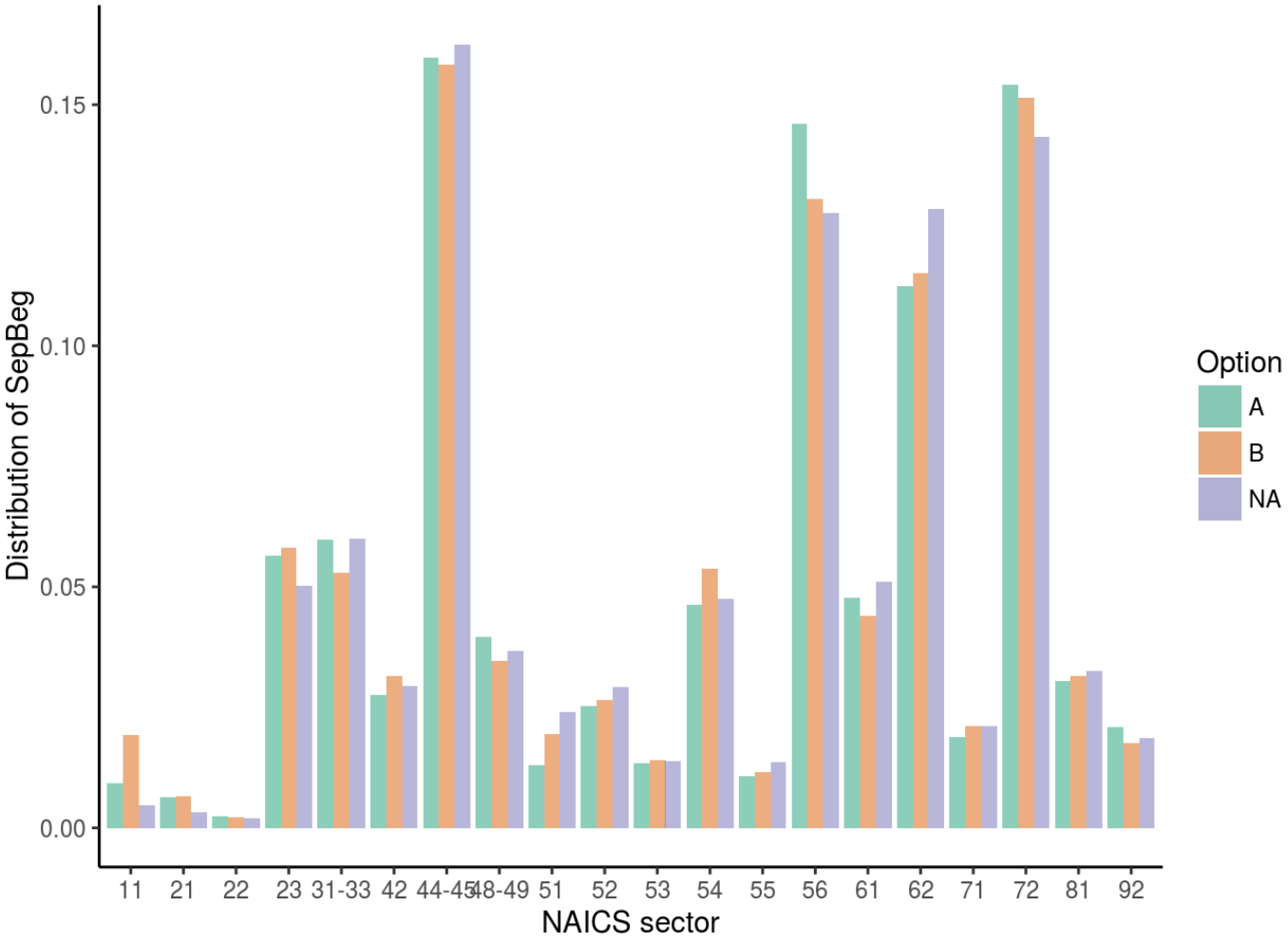
and merge on the indicators for MOU status:

```
allmous <- merge(all,mous,by.x="geography",by.y = "fips",all.x = TRUE)
size <- allmous[allmous$ind_level=="A",]
industry <- allmous[allmous$industry != "00",]
```

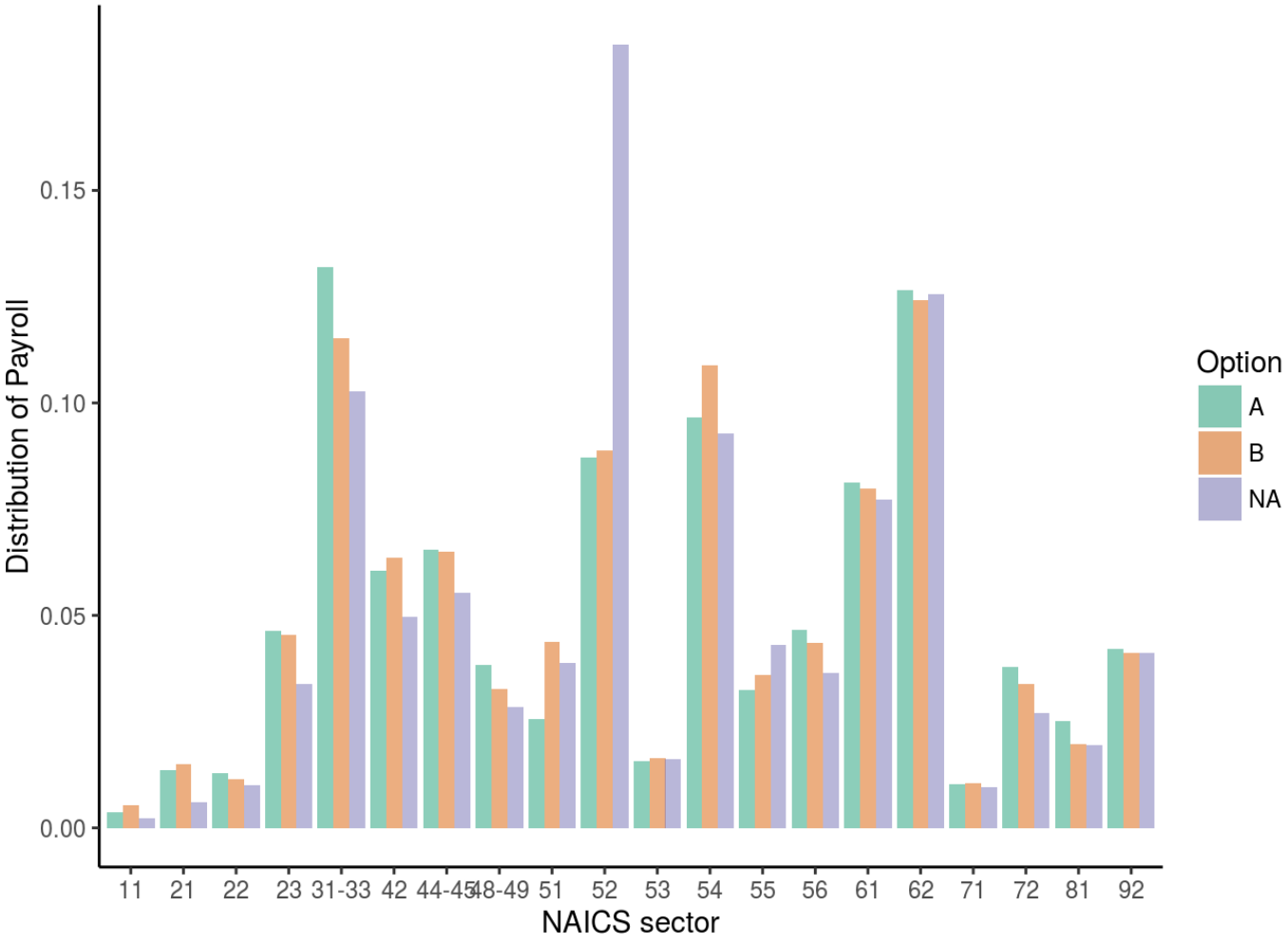
The industry distribution of **Emp** by chosen option thus looks like this:



For the industry distribution of **SepBeg**, the distribution looks like this:



For the industry distribution of **Payroll**, the distribution looks like this:



Additional variables (see the (LEHD%20Schema)http://lehd.ces.census.gov/data/schema/V4.0.4/lehd_public_use_schema.html (http://lehd.ces.census.gov/data/schema/V4.0.4/lehd_public_use_schema.html) for names) can be easily added to the Rmd source file.