Data Acquisition

Data from PDF table in paper

We extract out the table from Frey and Osborne's paper which if unfortunately in PDF form. We use the tabulizer library to extract it from the PDF. We trim off uneeded tables and then remove the header rows from each of the pages table. We then combine the paged tables into one table, and name the rows. We then filter out multi line rows by removing blank ranks. As a sanity check we can see that we have 702 rows in our table the same number of ranks in the paper.

```
dataTables<-extract_tables("https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employmen
trimTables<-dataTables[3:length(dataTables)]
trimTables<-purrr::map(trimTables,~ subset(as.data.frame(.x, stringsAsFactors=FALSE , row.names=NULL)))
trimTables<-purrr::map(trimTables, ~.x[3:nrow(.x),])
autoTable<-bind_rows(trimTables)[,1:5]
names(autoTable)<- dataTables[[3]][2,]
autoTable<-autoTable %>%  na_if("") %>% drop_na(Rank)
kable(head(autoTable))
```

Rank	Probability	Label	SOC code	Occupation
1.	0.0028	NA	29-1125	Recreational Therapists
2.	0.003	NA	49-1011	First-Line Supervisors of Mechanics, Installers, and Repairers
3.	0.003	NA	11-9161	Emergency Management Directors
4.	0.0031	NA	21-1023	Mental Health and Substance Abuse Social Workers
5.	0.0033	NA	29-1181	Audiologists
6.	0.0035	NA	29-1122	Occupational Therapists

Data from BLS

We can fairly directly download from BLS, though we do unfortunately have to pick out the right file of the two in the zip, which adds some complication.

```
destFile <-"state_M2018_dl.zip"
download.file("https://www.bls.gov/oes/special.requests/oesm18st.zip",destfile=destFile)
BLS18<-rio::import(unzip(destFile,"oesm18st/state_M2018_dl.xlsx"))
BLS18$JOBS_1000 <- (as.numeric(BLS18$JOBS_1000))
```

```
## Warning: NAs introduced by coercion
```

We can also grab the 09 data, which is unfortunately in a different storage format (prior to 09 it seems to be a differing data format)

```
destFile <-"state_M2009_dl.zip"
download.file("https://www.bls.gov/oes/special.requests/oesm09st.zip",destfile=destFile)
BLS09<-rio::import(unzip(destFile,"state_dl.xls"))
BLS09$JOBS_1000 <- (as.numeric(BLS09$JOBS_1000))
```

```
## Warning: NAs introduced by coercion
```

Analysis

We then simply take the Jobs per 1000 of the top and bottom job codes and sum them up per state. We also exclude DC as it is a small region that is unusual in job composition, and an outlier.

```
hardToAutomate<-head(autoTable$`SOC code`,100)
easyToAutomate<-tail(autoTable$`SOC code`,100)
stateLevelHard<-BLS18 %>% filter(ST!="DC") %>% group_by(STATE) %>% filter(OCC_CODE %in% hardToAutomate)
kable(head(stateLevelHard %>% arrange(n)))
```

STATE	n
Nevada	67.865
South Dakota	77.404
Guam	79.019
Florida	84.564
Wyoming	86.708
North Dakota	88.400

```
kable(head(stateLevelHard %>% arrange(desc(n))))
```

STATE	n
Massachusetts	130.231
Maryland	123.420
Connecticut	120.441
New York	112.725
Vermont	108.231
Arizona	105.557

```
stateLevelEasy<-BLS18 %>% filter(ST!="DC") %>% group_by(STATE) %>% filter(OCC_CODE %in% easyToAutomate)
kable(head(stateLevelEasy %>% arrange(n)))
```

STATE	n
Massachusetts	142.349
Virgin Islands	147.315
Maryland	150.661
Washington	153.513
New York	153.786
Michigan	155.326

```
kable(head(stateLevelEasy %>% arrange(desc(n))))
```

STATE	n
Nevada	185.500
Montana	183.932
New Hampshire	182.500
Florida	180.700

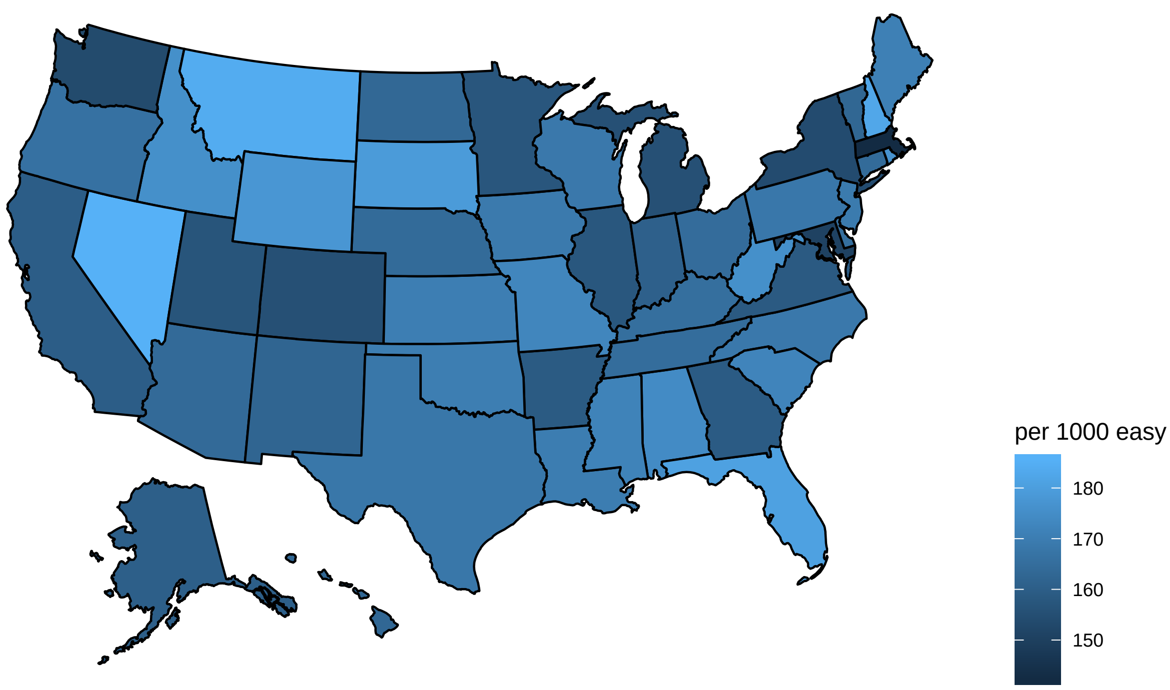
STATE	n
South Dakota	178.905
Puerto Rico	178.600

```
stateLevelHard09<-BLS09 %>% filter(ST!="DC") %>% group_by(STATE) %>% filter(OCC_CODE %in% hardToAutomate
stateLevelEasy09<-BLS09 %>% filter(ST!="DC") %>% group_by(STATE) %>% filter(OCC_CODE %in% easyToAutomate
kable(head(stateLevelHard09%>% arrange(n)))
```

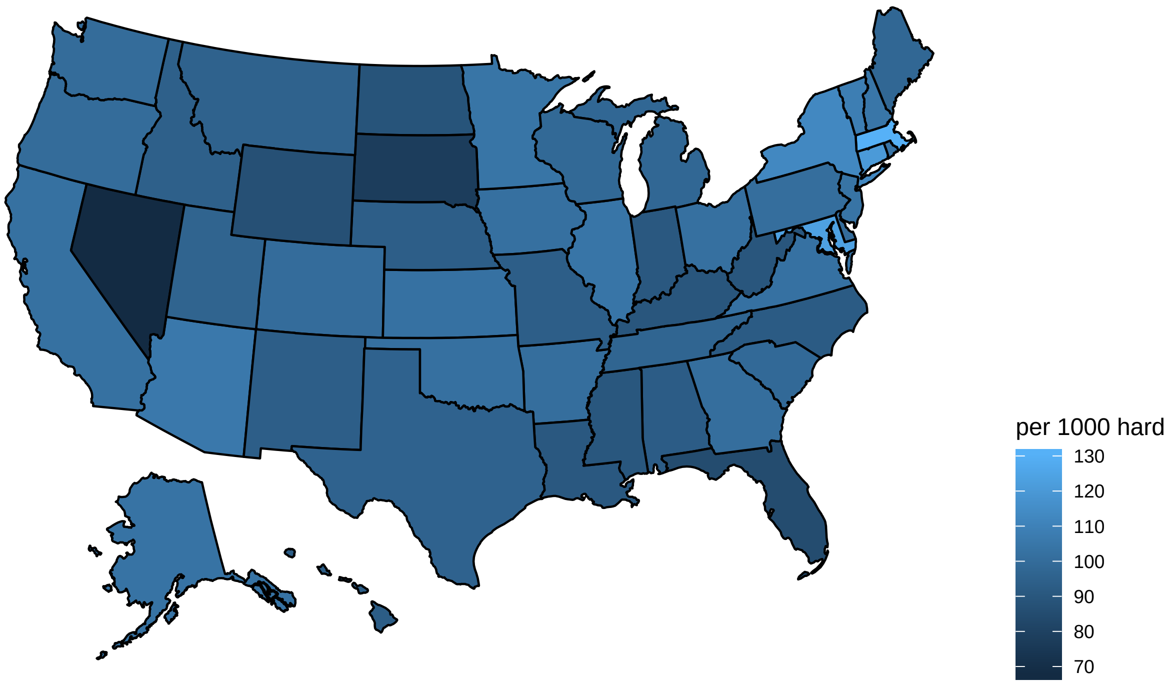
STATE	n
Puerto Rico	71.586
Guam	78.763
Nevada	80.378
Virgin Islands	91.813
Florida	96.083
South Dakota	97.161

Mapping

```
names(stateLevelEasy)<-tolower(names(stateLevelEasy))
names(stateLevelHard)<-tolower(names(stateLevelHard))
plot_usmap(data = stateLevelEasy, values = "n", color = "black") + scale_fill_continuous(name="per 1000
```



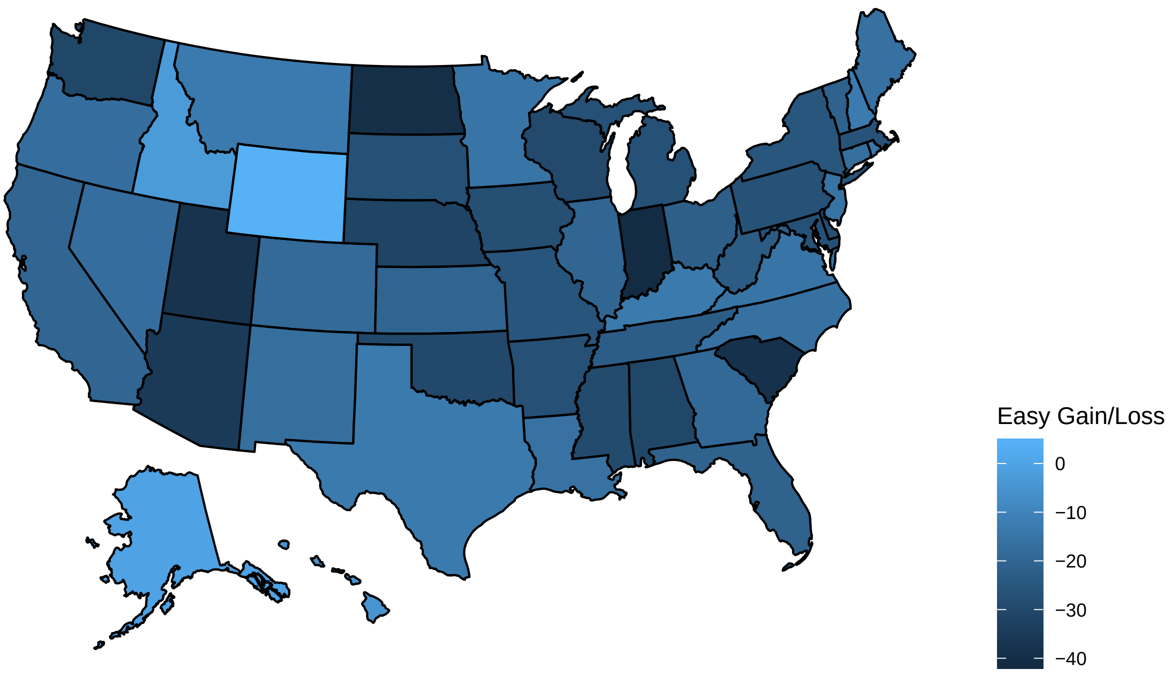
```r
plot_usmap(data = stateLevelHard, values = "n", color = "black") + scale_fill_continuous(name="per 1000
```



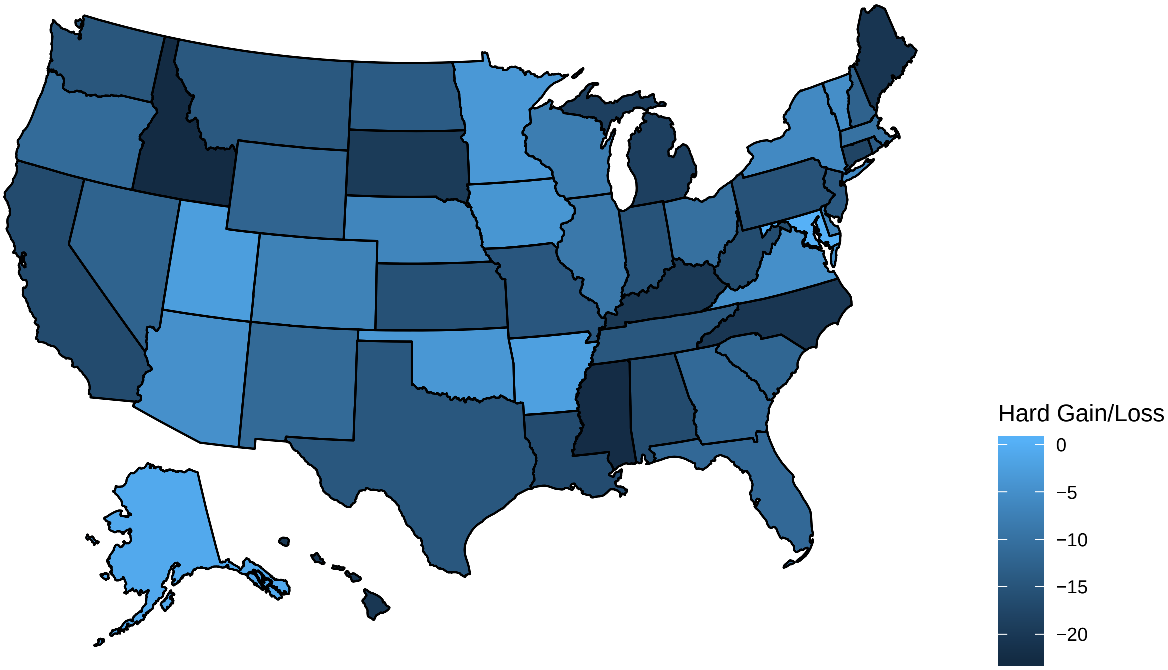
Looking at this we can see that Nevada looks to be in trouble with lots of easy to automate industry workers, and few hard.

We can go a bit deeper and look at the difference between 2009 and 2018

```r
names(stateLevelEasy09)<-c("state","priorN")
names(stateLevelHard09)<-c("state","priorN")
stateLevelEasyChange <- stateLevelEasy %>% inner_join (stateLevelEasy09,by=c("state"),name="priorN") %>%
plot_usmap(data=stateLevelEasyChange, values ="gain", color = "black") + scale_fill_continuous(name="Ea
```



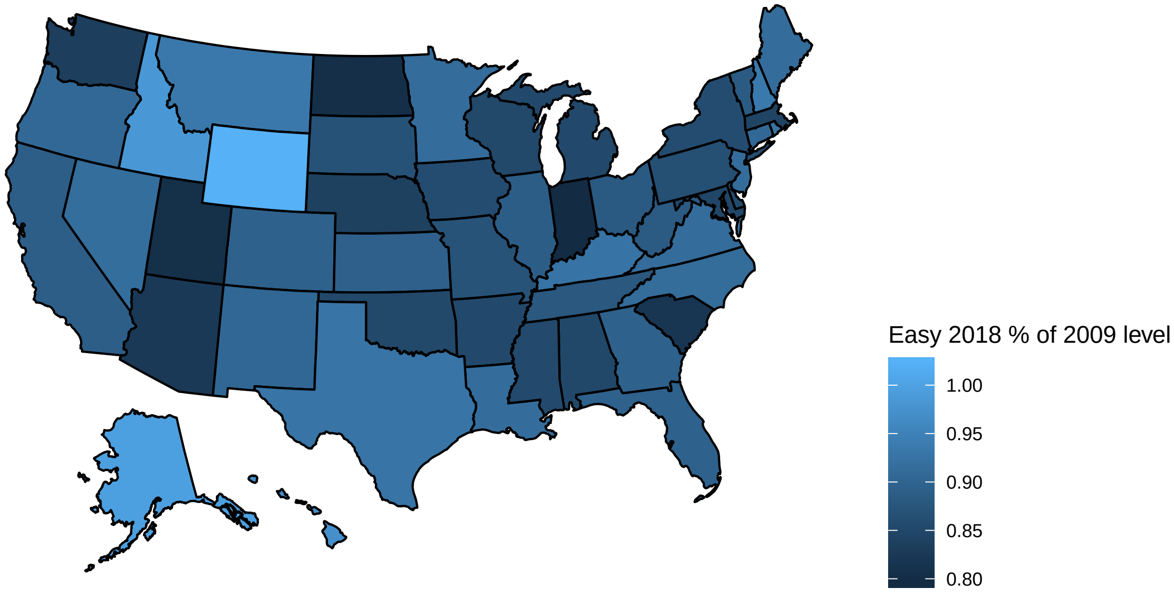

```
stateLevelHardChange <- stateLevelHard %>% inner_join (stateLevelHard09,by=c("state"),name="priorN") %>%
plot_usmap(data=stateLevelHardChange, values ="gain", color = "black") + scale_fill_continuous(name="Ha
```



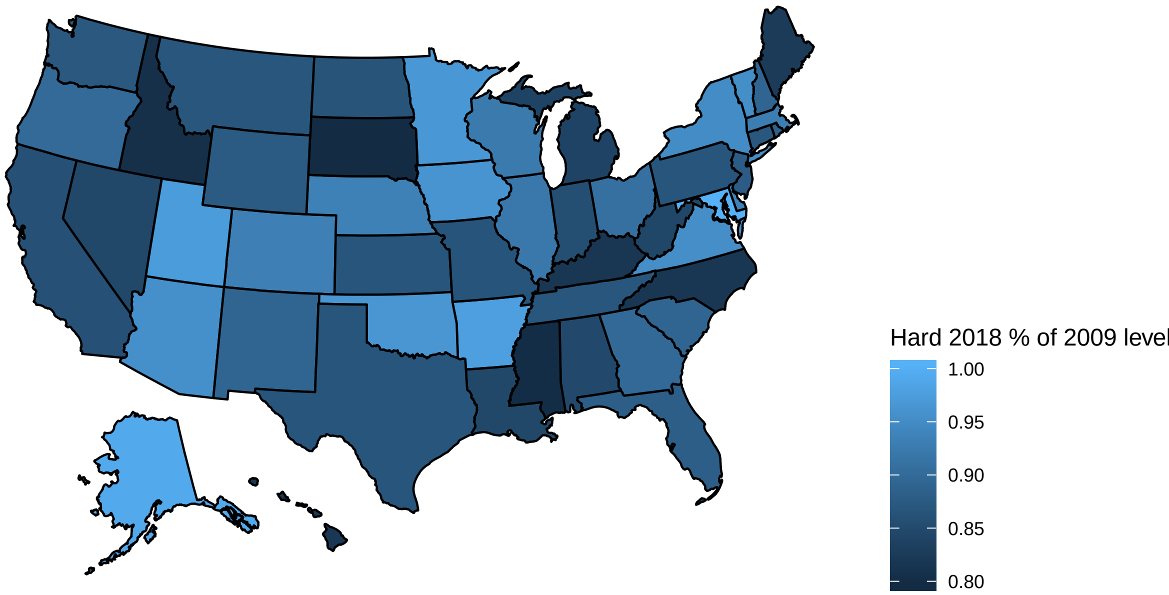

```
stateLevelEasyDelta <- stateLevelEasy %>% inner_join (stateLevelEasy09,by=c("state"),name="priorN") %>%
kable(head(stateLevelEasyDelta %>% arrange(gain)))
```

state	n	priorN	gain
Indiana	160.764	201.774	0.7967528
North Dakota	163.308	202.902	0.8048615
Utah	156.893	194.783	0.8054758
South Carolina	172.032	210.636	0.8167265
Arizona	163.967	199.148	0.8233424
Washington	153.513	184.234	0.8332501

```
plot_usmap(data=stateLevelEasyDelta, values ="gain", color = "black") + scale_fill_continuous(name="Eas
```




```
stateLevelHardDelta <- stateLevelHard %>% inner_join (stateLevelHard09,by=c("state"),name="priorN") %>%
plot_usmap(data=stateLevelHardDelta, values ="gain", color = "black") + scale_fill_continuous(name="Harc
```



Interestingly enough it seems both hard and easy to automate fields are losing numbers. This suggests a broadening that automation isn't the only factor in job loss.

We can briefly look at growing and falling occupations.

```
OCCData09<-BLS09 %>% group_by(`OCC_CODE`)  %>% tally(JOBS_1000) %>%select(OCC_CODE, n)
OCCData18<-BLS18 %>% group_by(`OCC_CODE`)  %>% tally(JOBS_1000) %>%select(OCC_CODE, n)
OCCDataDelta <- OCCData18 %>% inner_join (OCCData09,by=c("OCC_CODE"),name="priorN") %>% mutate(change=n
kable(head(OCCDataDelta%>% arrange(change),10))
```

OCC_TITLE	change
Locomotive Firers	0.0000000
Model Makers, Wood	0.0431655
Fabric Menders, Except Garment	0.0909091
Timing Device Assemblers and Adjusters	0.1090909
Entertainment Attendants and Related Workers, All Other	0.1186025
Patternmakers, Wood	0.1291866
Manufactured Building and Mobile Home Installers	0.1820470
Funeral Service Managers	0.2228232
Telephone Operators	0.2379895
Helpers–Roofers	0.2572308

```
kable(head(OCCDataDelta%>% arrange(desc(change)),10))
```

OCC_TITLE	change
Subway and Streetcar Operators	Inf
Radio, Cellular, and Tower Equipment Installers and Repairers	4.380591
Human Resources Specialists	2.825016
Personal Care Aides	2.666869
Social Sciences Teachers, Postsecondary, All Other	2.403361
Tire Builders	2.353550
Astronomers	2.329268
Airfield Operations Specialists	2.109386
Financial Examiners	2.059019
Manicurists and Pedicurists	2.048853

Looking at these, one can see that there are broader technological and sociological changes that aren't automation per se, such as a decline in telephone operators.