

```
1 # FiveThirtyEight.com
2 # Article: "Using 'Infrastructure Jobs' as a Measuring Stick For State-Level Spending"
3 # Published on: June 3, 2014
4 # Article Author: Andrew Flowers (andrew.flowers@fivethirtyeight.com)
5 # Article URL: http://fivethirtyeight.com/datalab/using-infrastructure-jobs-as-a-
  measuring-stick-for-state-level-spending/
6
7 # Code Author: Andrew Flowers (andrew.flowers@fivethirtyeight.com)
8 # Dependent files: payroll-states.csv
9
10 # Purpose: Get state-level data on "Heavy Construction and Civil Engineering"
11 # Will produce statepayrolls.csv file after running
12
13 # Get data
14 temp<-tempfile()
15 download.file("http://download.bls.gov/pub/time.series/sm/sm.data.62.Construction.Curren
  t",temp)
16 statepay.raw<-
  read.table(temp,header=TRUE,sep="\t",stringsAsFactors=FALSE,strip.white=TRUE)
17 unlink(temp)
18
19 # Add series info
20 series<-
  read.table("http://download.bls.gov/pub/time.series/sm/sm.series",sep="\t",header=TRUE,s
  trip.white=TRUE)
21 state<-read.csv("payroll-states.csv",header=TRUE,strip.white=TRUE)
22 series<-merge(series,state,by="state_code")
23
24 # Add industry info
25 industry<-read.table("http://download.bls.gov/pub/time.series/sm/sm.industry", sep="\t",
  header=TRUE, strip.white=TRUE)
26 industry$industry_name<-NULL
27 industry$industry_name<-row.names(industry)
28 row.names(industry)<-NULL
29 names(industry)<-c("industry_name","industry_code")
30
31 series<-merge(series,industry,by="industry_code")
32
33 statepay<-merge(statepay.raw,series,by="series_id")
34
35 # Take out heavy construction industry data (which is coded 20237000)
36 heavyIndCodes<-c(20237000, 20237100, 20237200, 20237300, 20237900)
37 statepay.heavy<-statepay[greps(heavyIndCodes[1], statepay$industry_code),]
38
39 # Clean state data
40 statepay.NSA<-subset(statepay.heavy,!period=="M13")
41 statepay.NSA<-subset(statepay.NSA, area_code==0)
42 statepay.NSA$date<-as.Date(paste(statepay.NSA$year,statepay.NSA$period,"01",sep="-
  "),"%Y-M%m-%d")
43 statepay.NSA<-subset(statepay.NSA,select=c("series_id","date","state_name","value"))
44
45 # Convert to time series
46 require(reshape2)
47 statepay.NSA.t<-dcast(statepay.NSA, date ~ state_name,value.var="value") #
  ,fun.aggregate=mean)
48 write.csv(statepay.NSA.t,file="statepayrolls.csv")
```

49

50

51

