tinytex::install_tinytex()

title: “Bhasin-S-hw1-1” output: pdf_document date: “2023-01-23” —

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

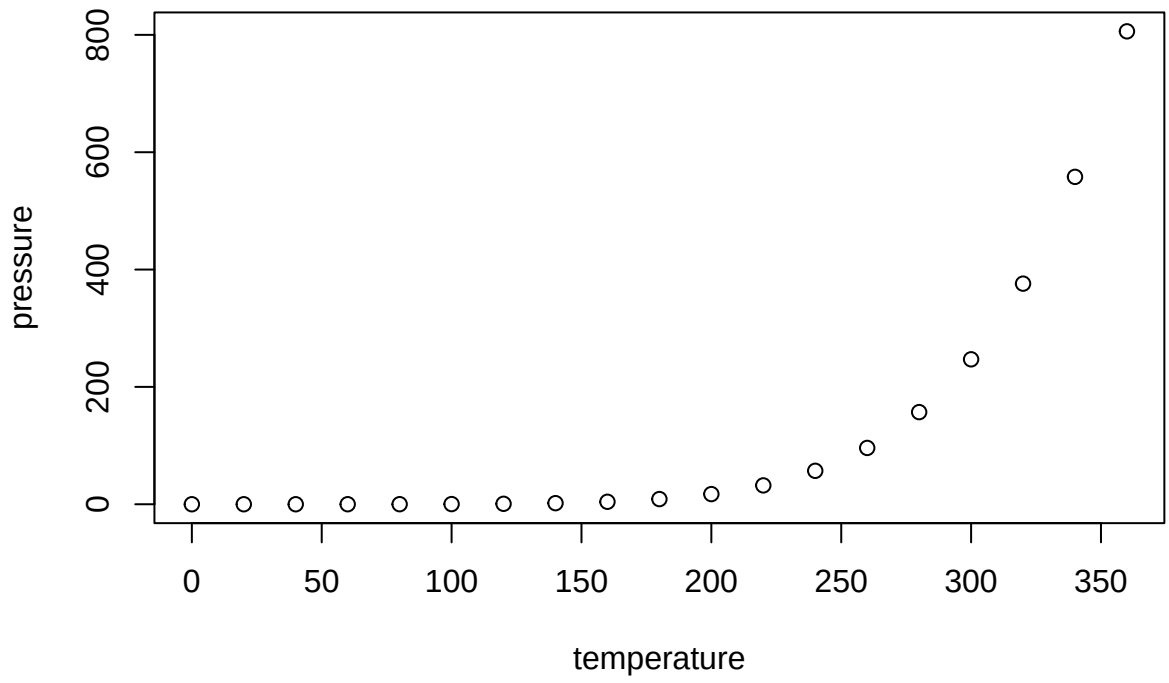
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Read in enrollment data for january of each year

install.packages("usethis") install.packages("tidyverse") library(tidyverse)

for (y in 2007:2015) { ## Basic contract/plan information ma.path=paste0("data/input/monthly-ma-and-pdp-enrollment-by-cpsc/CPSC_Contract_Info_",y,"_01.csv") contract.info=read_csv(ma.path, skip=1, col_names = c("contractid","planid","org_type","plan_type", "partd","snp","eghp","org_name","org_marketing_name", "plan_name","parent_org","contract_date"), col_types = cols(contractid = col_character(), planid = col_double(), org_type = col_character(), plan_type = col_character(), partd = col_character(), snp = col_character(), eghp = col_character(), org_name = col_character(), org_marketing_name = col_character(), plan_name = col_character(), parent_org = col_character(), contract_date = col_character()))

contract.info = contract.info %>% group_by(contractid, planid) %>% mutate(id_count=row_number())

contract.info = contract.info %>% filter(id_count==1) %>% select(-id_count)

Enrollments per plan ma.path=paste0("data/input/monthly-ma-and-pdp-enrollment-by-cpsc/CPSC_Enrollment_Info_ enroll.info=read_csv(ma.path, skip=1, col_names = c("contractid","planid","ssa","fips","state","county","enrollment"), col_types = cols(contractid = col_character(), planid = col_double(), ssa = col_double(), fips = col_double(), state = col_character(), county = col_character(), enrollment = col_double()),na="*")

Merge contract info with enrollment info plan.data = contract.info %>% left_join(enroll.info, by=c("contractid", "planid")) %>% mutate(year=y)

Fill in missing fips codes (by state and county) plan.data = plan.data %>% group_by(state, county) %>% fill(fips)

Fill in missing plan characteristics by contract and plan id plan.data = plan.data %>% group_by(contractid, planid) %>% fill(plan_type, partd, snp, eghp, plan_name)

Fill in missing contract characteristics by contractid plan.data = plan.data %>% group_by(contractid) %>% fill(org_type,org_name,org_marketing_name,parent_org)

Collapse from monthly data to yearly plan.year = plan.data %>% group_by(contractid, planid, fips) %>% arrange(contractid, planid, fips) %>% rename(avg_enrollment=enrollment)

write_rds(plan.year,paste0("data/output/ma_data_",y,".rds")) }

full.ma.data <- read_rds("data/output/ma_data_2007.rds") for (y in 2008:2015) { full.ma.data <- rbind(full.ma.data,read_rds(paste0("data/output/ma_data_",y,".rds"))) }

write_rds(full.ma.data,"data/output/full_ma_data.rds") sapply(paste0("ma_data_", 2007:2015, ".rds"), unlink)

#Homework 1

#Enrollment Data

#1. There are 19,126,783 observations in my current data set.

full.ma.data %>% count(plan_type)

#2 There are 5,847,057 different plan_type in the data

file.path(plan.data)

plan.data <- read.csv("plan.data.csv") plan_type <- plan.data$plan_type

knitr::kable(plan.type, col.names=c("2010","2011","2012","2013","2014","2015"), type="html", caption = "Plan Count by Year", booktabs = TRUE)

#3

full.ma.data <- readRDS('data/full.ma.data.rds')