PS2

Alex Stephenson

2023-02-13

Part 1

```
vec1 = 1:1000
set.seed(12345)
vec2 = sample(vec1)
dat = data.frame(vec1, vec2)
head(dat)
```

```
##   vec1 vec2
## 1    1  142
## 2    2   51
## 3    3  720
## 4    4  730
## 5    5  220
## 6    6  664
```

```
idx = which(dat$vec2 %in% c(2, 47, 290, 812))
dat$vec2[idx] = NA
names(dat) = c("caseid", "wage")

funcs = function(x){
  c(mean = mean(x, na.rm = T),
    med = median(x, na.rm = T),
    std = sd(x, na.rm = T))
}

sapply(dat, funcs)[,2]
```

```
##     mean      med      std
## 501.3544 501.5000 288.3622
```

```
dat2 = na.omit(dat)
head(dat2)
```

```
##   caseid wage
## 1      1  142
## 2      2   51
## 3      3  720
## 4      4  730
## 5      5  220
## 6      6  664
```

Part 2

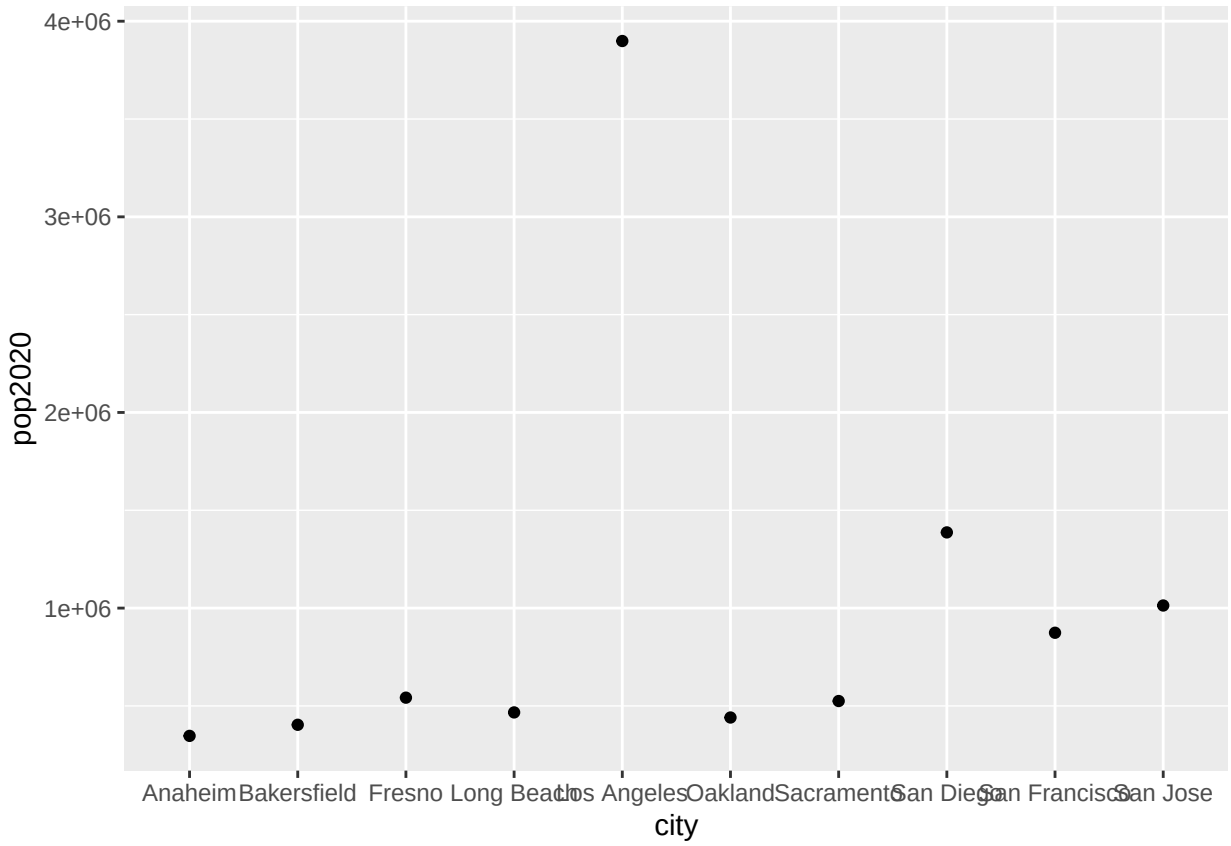
```
cities = read.csv("CAcities.csv")

for(i in 1:nrow(cities)){
  print(cities$city[i])
}
```

```
## [1] "Anaheim"
## [1] "Bakersfield"
## [1] "Fresno"
## [1] "Long Beach"
## [1] "Los Angeles"
## [1] "Oakland"
## [1] "Sacramento"
## [1] "San Diego"
## [1] "San Francisco"
## [1] "San Jose"
```

```
cities2 = cities[order(cities$pop2020,decreasing = T),]
for(i in 1:nrow(cities2)){
  print(cities2$city[i])
}
```

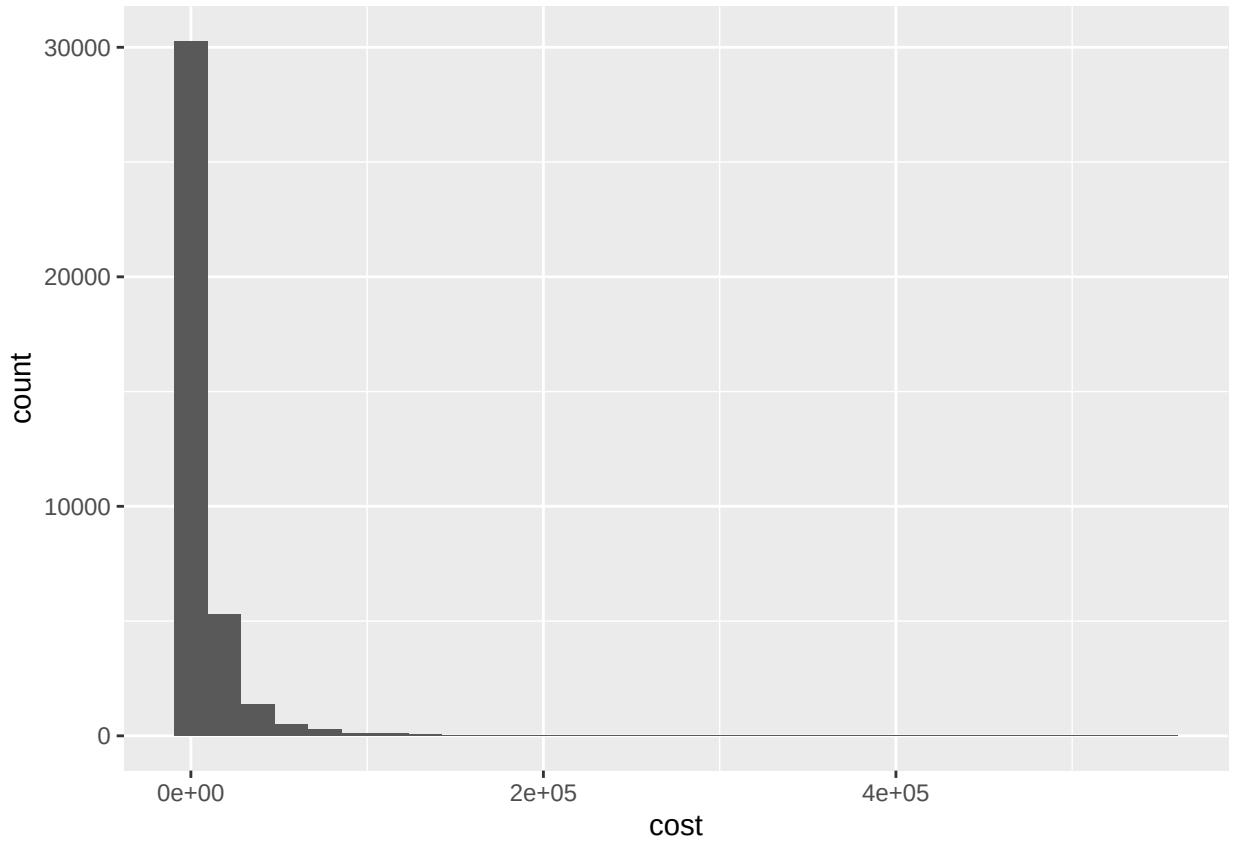
```
## [1] "Los Angeles"
## [1] "San Diego"
## [1] "San Jose"
## [1] "San Francisco"
## [1] "Fresno"
## [1] "Sacramento"
## [1] "Long Beach"
## [1] "Oakland"
## [1] "Bakersfield"
## [1] "Anaheim"
```

```
cities |>
  ggplot(aes(x=city, y = pop2020))+
  geom_point()
```

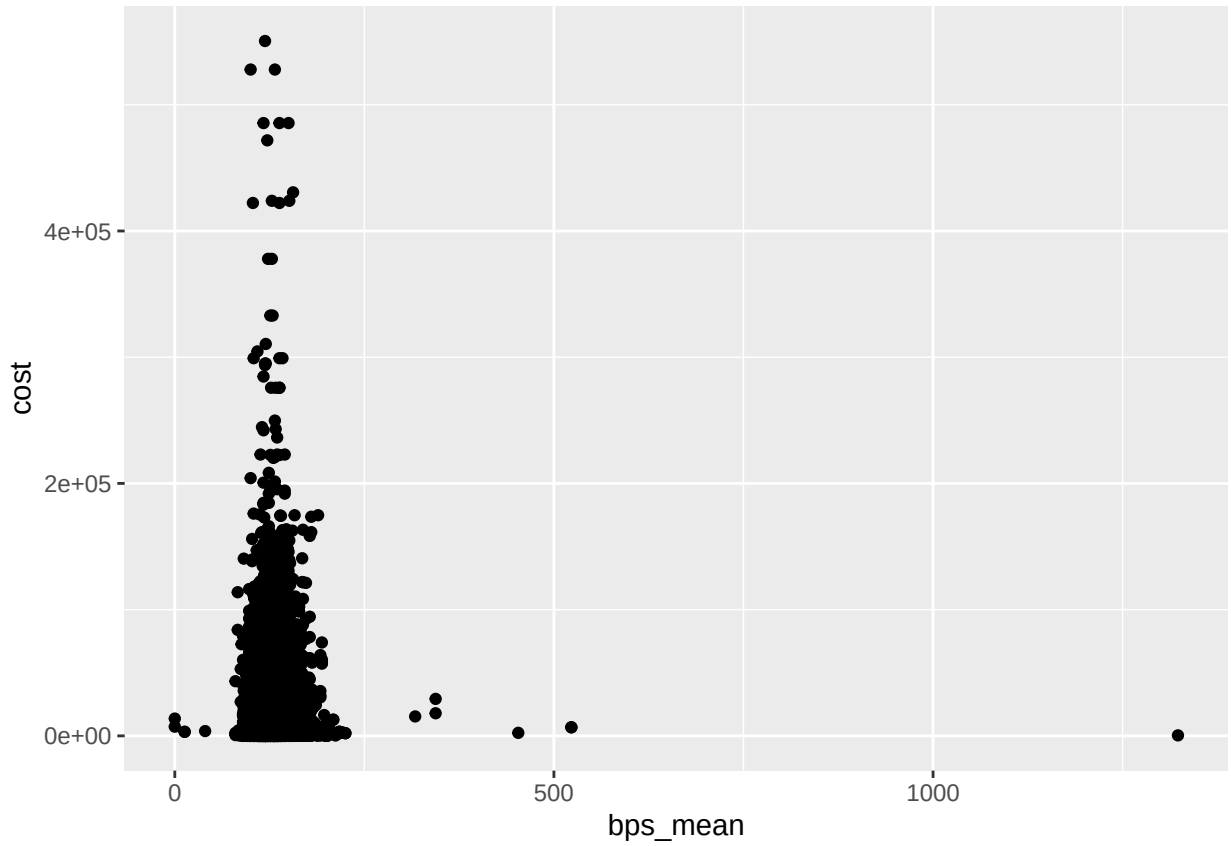


Part 3

```
hdat = read.csv("data_health_synth_small.csv") |>
  na.omit()

hdat |>
  ggplot(aes(x = cost))+
  geom_histogram()
```




```
hdat |>
  ggplot(aes(x = bps_mean, y = cost))+
  geom_point()
```




```
set.seed(12345)
cost_samp = sample(hdat$cost, replace = T)

mean(cost_samp)
```

```
## [1] 8524.394
```

```
mean(hdat$cost)
```

```
## [1] 8634.66
```

```
N = 1000
costs = vector(mode = "logical", length = N)
set.seed(12345)
for(i in 1:N){
  costs[i] = mean(sample(hdat$cost, replace = T))
}
sd(costs)
```

```
## [1] 95.84418
```

```
my_samps_function = function(x){
  ## x is a vector
  out = vector(mode = "logical", length = 1000)
  for(i in 1:1000){
    out[i] = mean(sample(x, replace = T))
  }
  return(sd(out))
}
```

```
set.seed(12345)
my_samps_function(hdat$cost)
```

```
## [1] 95.84418
```

```
set.seed(12345)
my_samps_function(hdat$bps_mean)
```

```
## [1] 0.08285895
```