Spatial Statistics in R

Venn Datagram

2024-07-17

Data Import and Initial Look

Link to Data

```r
library(dplyr)
```

```
##
## Attaching package: ’dplyr’
```

```
## The following objects are masked from ’package:stats’:
##
##     filter, lag
```

```
## The following objects are masked from ’package:base’:
##
##     intersect, setdiff, setequal, union
```

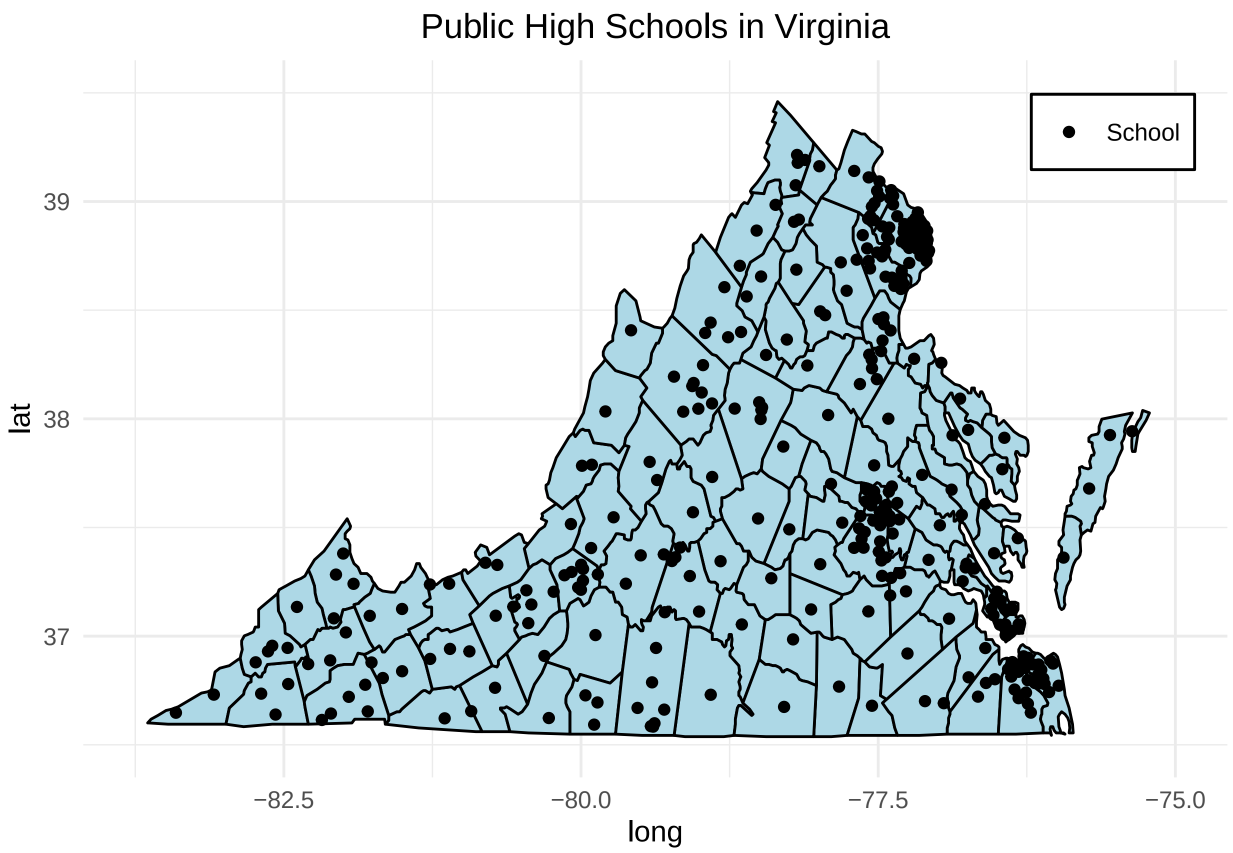
```r
df <- read.csv(
  "Public_School_Characteristics_-_Current.csv",
  encoding = "UTF-8")
```

```r
t.df <- df %>% filter(
  STABR == "VA",
  is.na(TOTAL) == FALSE,
  SCHOOL_LEVEL=="High") %>% select(
    LEA_NAME,SCH_NAME,LSTREET1,
    LSTREET2,LCITY,LZIP,
    LZIP4,ULOCALE,NMCNTY,
    TOTAL,LATCOD,LONCOD)
head(t.df)
```

```
##                                           LEA_NAME
## 1 Virginia School for the Deaf and Blind-Staunton
## 2                  Accomack County Public Schools
## 3                  Accomack County Public Schools
## 4                  Accomack County Public Schools
## 5                 Albemarle County Public Schools
## 6                 Albemarle County Public Schools
##                                       SCH_NAME               LSTREET1 LSTREET2
## 1 Virginia School for the Deaf and Blind High           104 VSDB Dr.
## 2                           CHINCOTEAGUE HIGH           4586 Main St
## 3                                 NANDUA HIGH 26350 Lankford Highway
## 4                                ARCADIA HIGH  8210 Lankford Highway
## 5                              ALBEMARLE HIGH    2775 Hydraulic Road
## 6                        COMMUNITY LAB SCHOOL     1200 Forest Street
##              LCITY  LZIP LZIP4           ULOCALE               NMCNTY TOTAL
## 1        Staunton 24401    NA    13-City: Small        Staunton city    40
## 2    Chincoteague 23336    NA  43-Rural: Remote      Accomack County   295
## 3           Onley 23418    NA  43-Rural: Remote      Accomack County   636
## 4        Oak Hall 23416    NA 42-Rural: Distant      Accomack County   713
## 5 Charlottesville 22901  8917  41-Rural: Fringe     Albemarle County  1987
## 6 Charlottesville 22903  5262    13-City: Small Charlottesville city   192
##     LATCOD    LONCOD
## 1 38.15066 -79.06397
## 2 37.94280 -75.36380
## 3 37.67920 -75.72560
## 4 37.92530 -75.55010
## 5 38.07620 -78.50120
## 6 38.04071 -78.48276
```

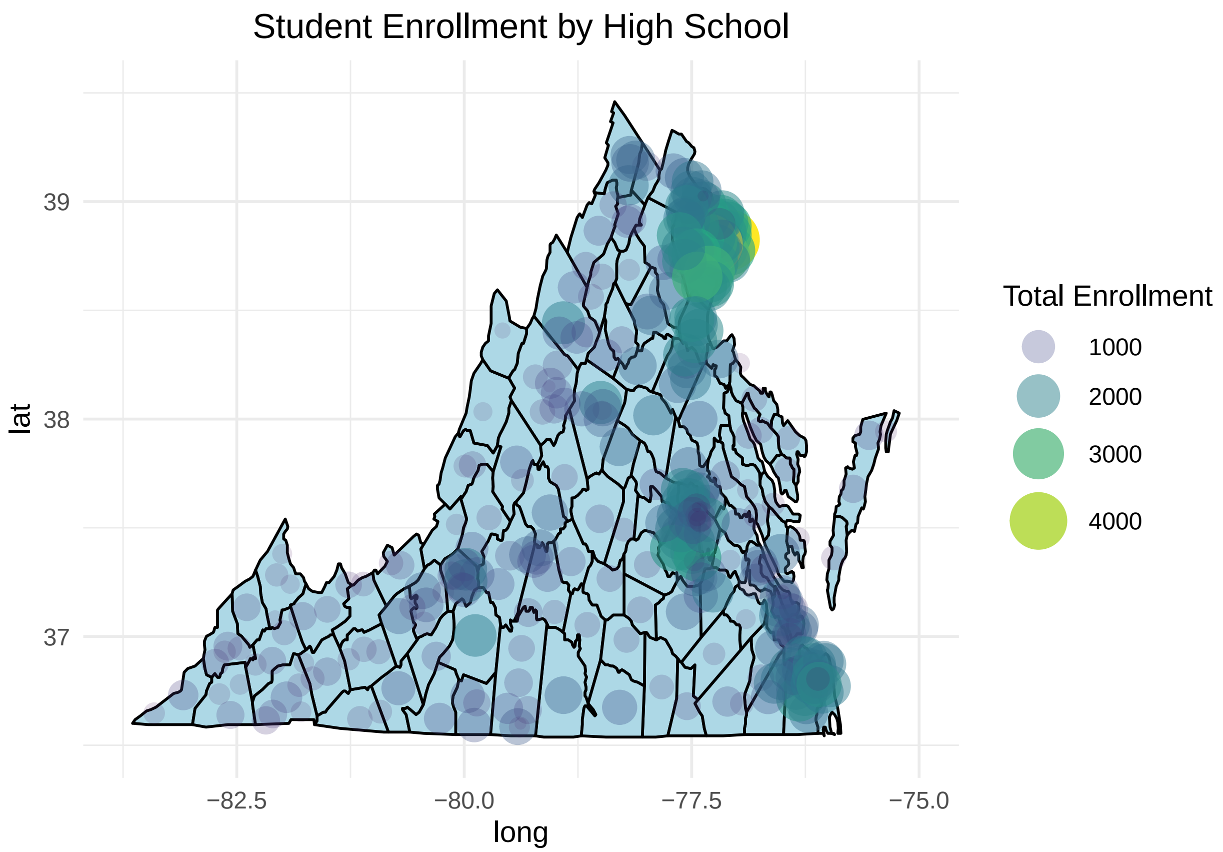
State Map

```
library(ggplot2)
virginia <- map_data('county', region = 'virginia')

ggplot(virginia,aes(x = long, y = lat, group = group)) +
  geom_polygon(fill = "lightblue", color = "black") +
  geom_point(
    data = t.df,
    aes(group = NULL, x = LONCOD, y = LATCOD, color = "School")
    ) +
  scale_color_manual(values = "black", name = "") +
  coord_cartesian(xlim = c(-83.75, -75), ylim = c(36.5, 39.5)) +
  theme_minimal() +
  labs(title = "Public High Schools in Virginia") +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = c(0.9, 0.9),
    legend.background = element_rect(
      fill = "white", color = "black", size = 0.5),
    legend.box.background = element_rect(
      fill = "white", color = "black", size = 0.5),
    legend.title = element_blank()
  )
```



Weights

```
ggplot(virginia, aes(x = long, y = lat, group = group)) +
  geom_polygon(fill = "lightblue", color = "black") +
  geom_point(
    data = t.df,
    aes(
      group=NULL, x = LONCOD, y = LATCOD,
      size = TOTAL, color = TOTAL,alpha=TOTAL)) +
  coord_cartesian(xlim = c(-83.75, -75), ylim = c(36.5, 39.5)) +
  scale_color_viridis_c(name = "Total Enrollment") +
  scale_alpha_continuous(name = "Total Enrollment") +
  scale_size_continuous(name = "Total Enrollment", range = c(1, 10)) +
  guides(
    color = guide_legend(), size = guide_legend(),
    alpha = guide_legend()) +
  theme_minimal() +
  labs(title = "Student Enrollment by High School") +
  theme(plot.title = element_text(hjust = 0.5))
```



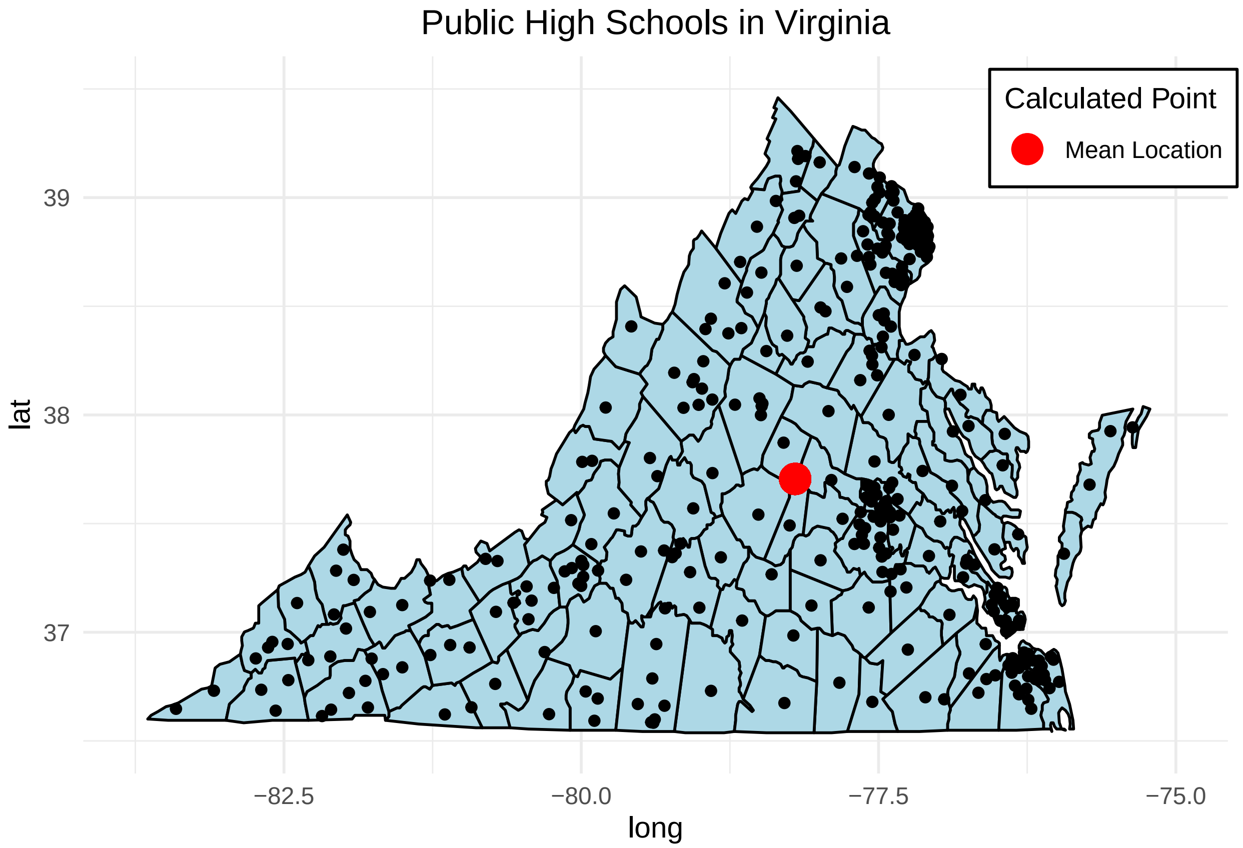
Mean and Median Calculations

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i}^{n} w_i}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i}^{n} w_i}$$

Equally Weighted Mean

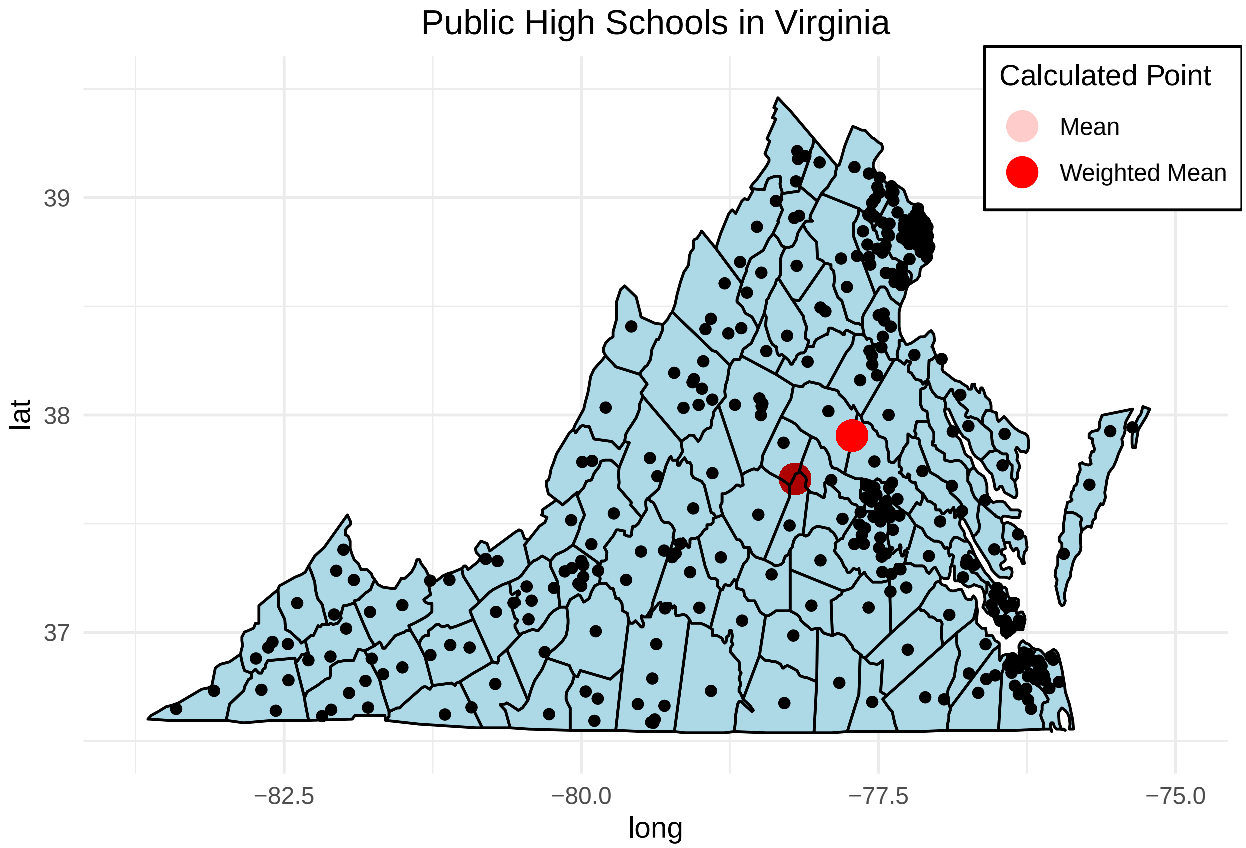
```
#Equally Weighted
mn.x <- mean(t.df$LONCOD)
mn.y <- mean(t.df$LATCOD)

ggplot(virginia, aes(x = long, y = lat, group = group)) +
  geom_polygon(fill = "lightblue", color = "black") +
  geom_point(data=t.df, aes(group=NULL,x=LONCOD,y=LATCOD)) +
  geom_point(aes(x=mn.x,y=mn.y,color='Mean Location'),size=5) +
  scale_color_discrete(type = 'red') +
  coord_cartesian(xlim =c(-83.75,-75), ylim = c(36.5,39.5)) +
  theme_minimal() +
  labs(
    title = "Public High Schools in Virginia",
    colour = "Calculated Point") +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = c(0.9, 0.9),
    legend.background = element_rect(
      fill = "white", color = "black", size = 0.5),
    legend.box.background = element_rect(
      fill = "white", color = "black", size = 0.5)
    )
```



Weighted Mean

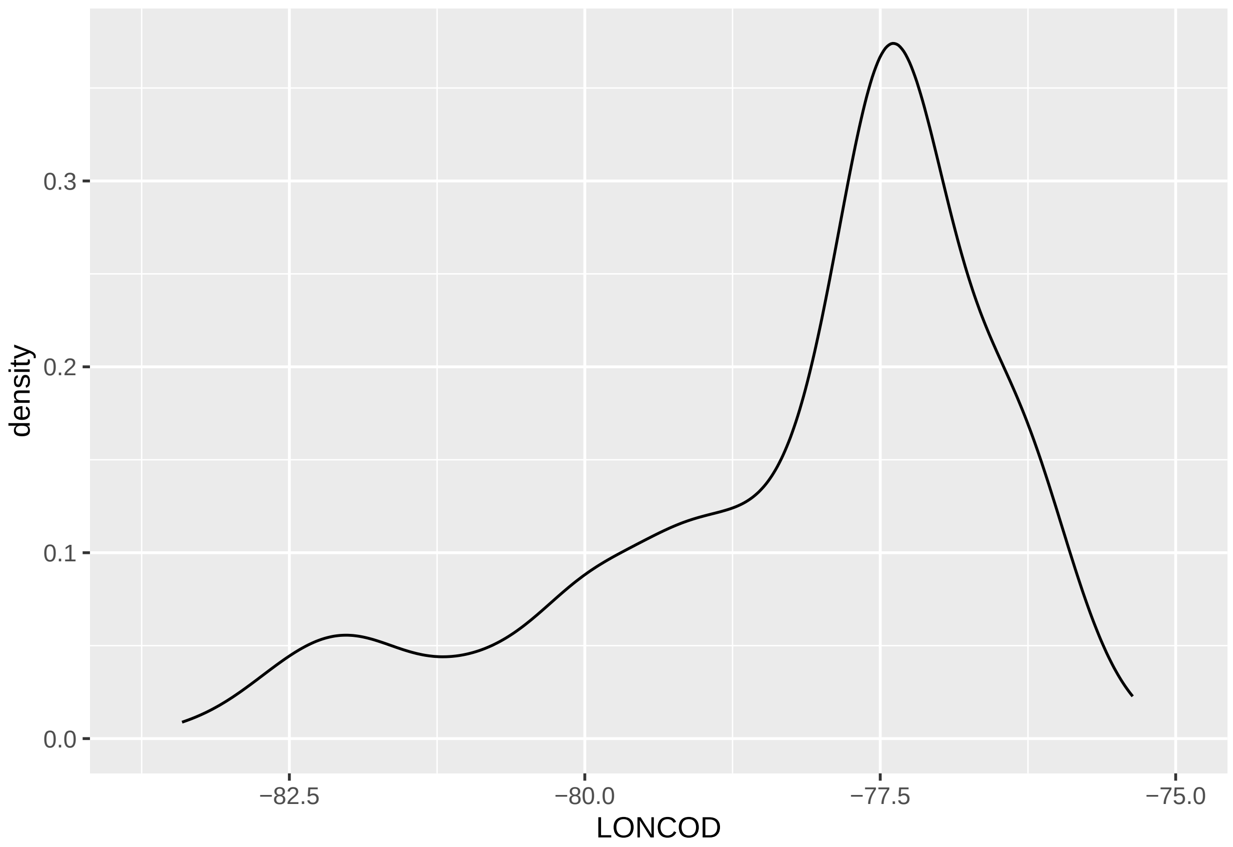
```
#Calculate Weighted Mean
n.total <- sum(t.df$TOTAL)
t.df$wts <- t.df$TOTAL / n.total

wtd.mn.x <- sum(t.df$wts * t.df$LONCOD)
wtd.mn.y <- sum(t.df$wts * t.df$LATCOD)
```

```
ggplot(virginia, aes(x = long, y = lat, group = group)) +
  geom_polygon(fill = "lightblue", color = "black") +
  geom_point(
    data=t.df, aes(group=NULL,x=LONCOD,y=LATCOD)) +
  geom_point(
    aes(x=mn.x,y=mn.y,colour="Mean"),size=5,alpha=0.002) +
  geom_point(
    aes(x=wtd.mn.x,y=wtd.mn.y,colour="Weighted Mean"),size=5) +
  scale_colour_discrete(type = c('red','red')) +
  coord_cartesian(
    xlim =c(-83.75,-75), ylim = c(36.5,39.5)) +
  theme_minimal() +
  labs(
    title = "Public High Schools in Virginia",
    colour = "Calculated Point") +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = c(0.9, 0.9),
    legend.background = element_rect(
      fill = "white", color = "black", size = 0.5),
    legend.box.background = element_rect(
      fill = "white", color = "black", size = 0.5)
    ) +
  guides(
    colour = guide_legend(override.aes = list(alpha = c(0.2, 1)))
  )
```

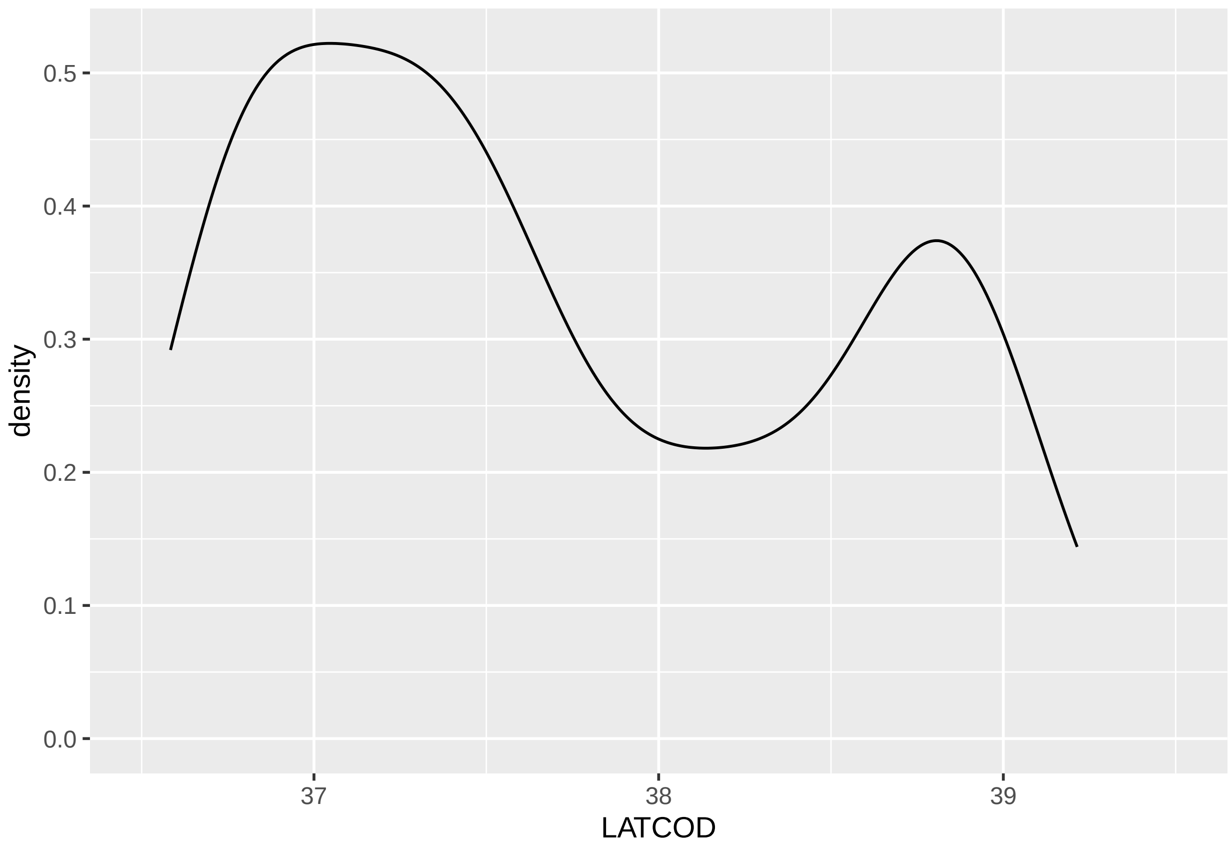


Distribution of Coordinates

```
ggplot(data=t.df, aes(x=LONCOD)) +
  geom_density() +
  coord_cartesian(xlim =c(-83.75,-75))
```



```
ggplot(data=t.df, aes(x=LATCOD)) +
  geom_density() +
  coord_cartesian(xlim =c(36.5,39.5))
```



Median

$$x' = \frac{\sum_{i=1}^{n} \frac{w_i x_i}{d_i}}{\sum_{i=1}^{n} \frac{w_i}{d_i}}$$

$$y' = \frac{\sum_{i=1}^{n} \frac{w_i y_i}{d_i}}{\sum_{i=1}^{n} \frac{w_i}{d_i}}$$

d_i = distance between point i and initial location.

```
#If not working with geospatial data
get.euc.dist <- function(start.x,start.y,lon,lat){
  return(sqrt((start.x-lon)^2 + (start.y-lat)^2))
}
```

```
#Use this formula
dists <- get.euc.dist(wtd.mn.x,wtd.mn.y,lon = t.df$LONCOD,lat=t.df$LATCOD)
```

```
#Full Process
init.pt.x <- wtd.mn.x
init.pt.y <- wtd.mn.y

coords <- data.frame(iteration = integer(), x = numeric(), y = numeric())

for (i in seq(100)){
  dists <- get.euc.dist(init.pt.x,init.pt.y,lon = t.df$LONCOD,lat=t.df$LATCOD)
  init.pt.x <- sum( (t.df$wts * t.df$LONCOD) / dists ) / sum( t.df$wts / dists)
  init.pt.y <- sum( (t.df$wts * t.df$LATCOD) / dists ) / sum( t.df$wts / dists)

  coords <- rbind(
    coords,
    data.frame(iteration = i, x = init.pt.x, y = init.pt.y))
}
c(init.pt.x,init.pt.y)
```

```
## [1] -77.50336  38.21054
```

Distance Formula for Median

$$d = 2r\sin^{-1}\left(\sqrt{\sin^2\left(\frac{\Phi_2-\Phi_1}{2}\right)+\cos(\Phi_1)\cos(\Phi_2)\sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right)$$

Φ_n= Latitude at n (think of latitude like fatitude, or (ph)atitutde like (ph)i)

λ_n = Longitude at n (longitude, think long limbs, long lambda)

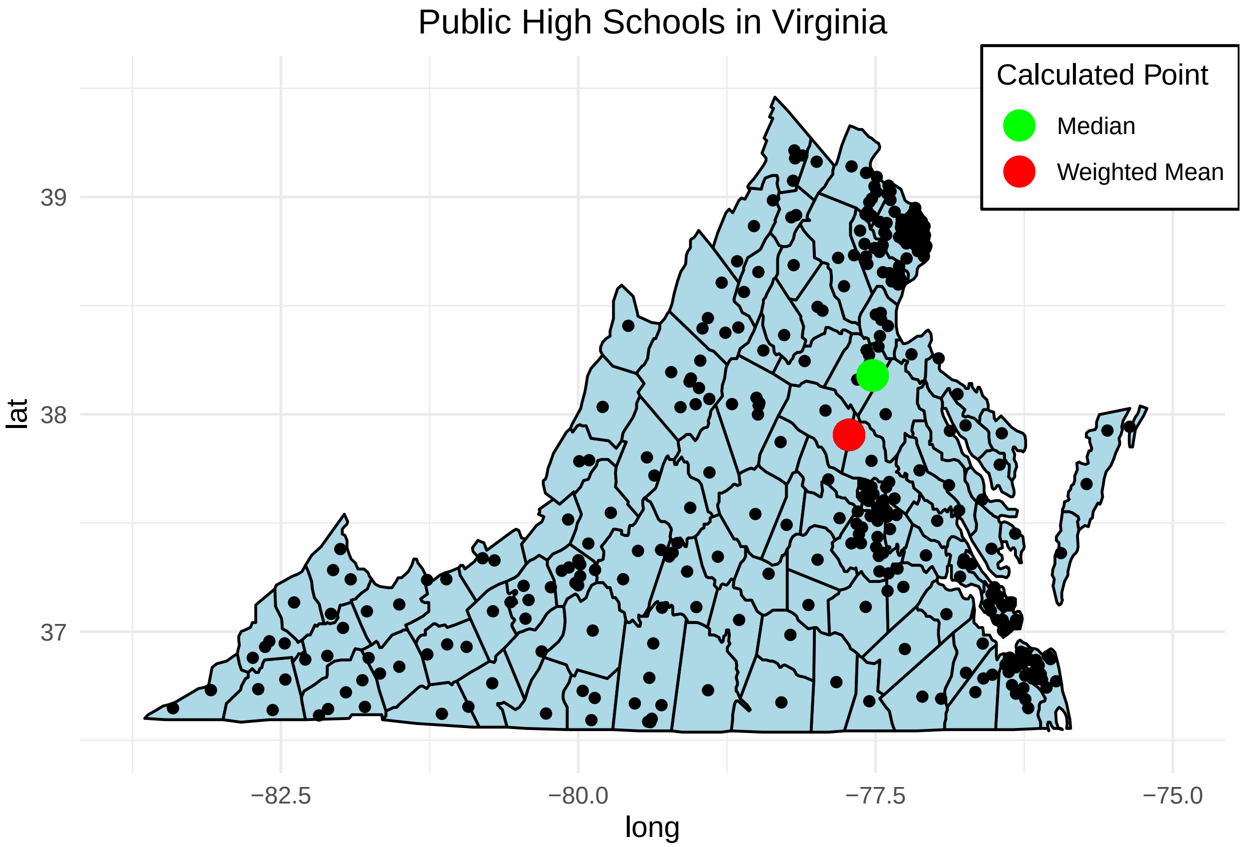
r = Radius (in this case we use Earth's so r = 6378137 m)

```
library(geosphere)

init.pt.x <- wtd.mn.x
init.pt.y <- wtd.mn.y

coords <- data.frame(iteration = integer(), x = numeric(), y = numeric())

for (i in seq(100)){
  dists <- distHaversine(cbind(t.df$LONCOD,t.df$LATCOD),c(init.pt.x,init.pt.y),)
  init.pt.x <- sum( (t.df$wts * t.df$LONCOD) / dists ) / sum( t.df$wts / dists)
  init.pt.y <- sum( (t.df$wts * t.df$LATCOD) / dists ) / sum( t.df$wts / dists)

  coords <- rbind(
    coords,
    data.frame(iteration = i, x = init.pt.x, y = init.pt.y))
}
c(init.pt.x,init.pt.y)
```

```
## [1] -77.52434  38.17895
```

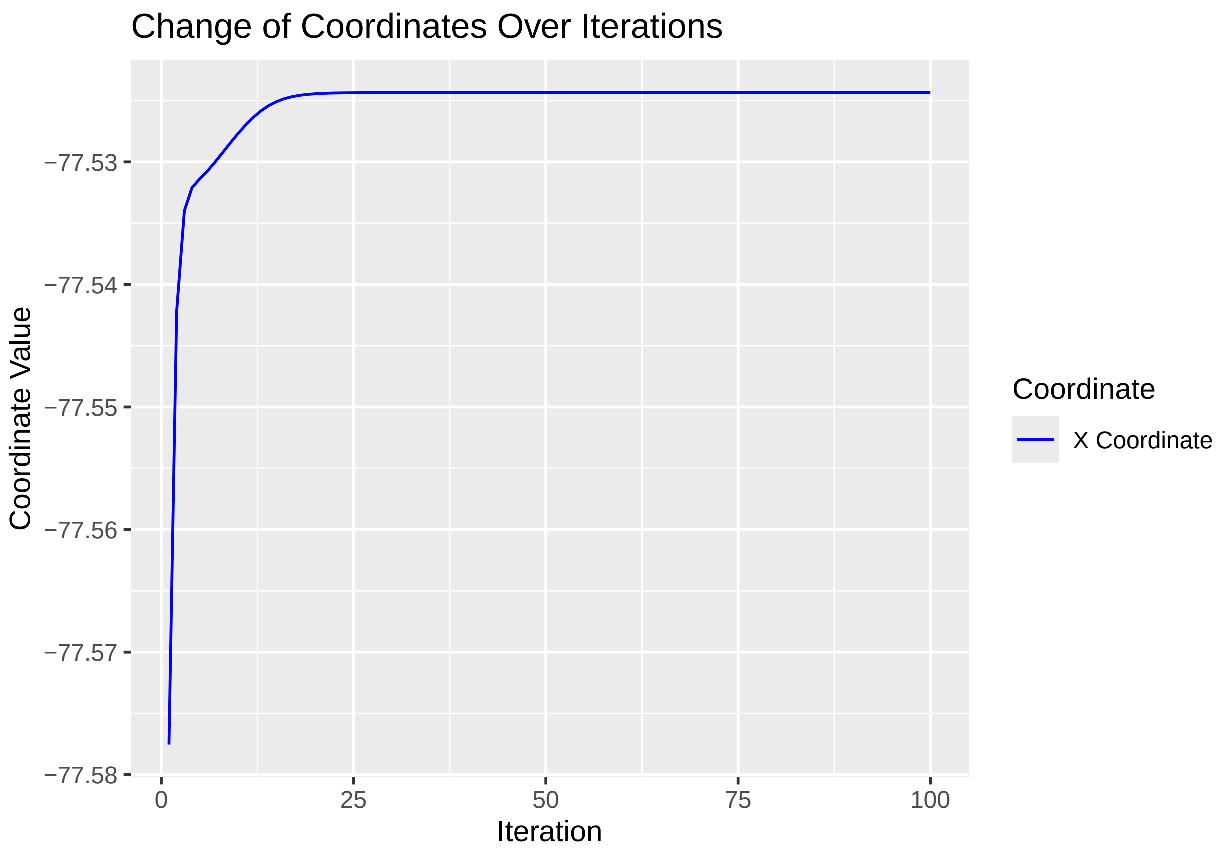
```
#Median Plot
points_to_label <- data.frame(
  x = c(wtd.mn.x, init.pt.x),
  y = c(wtd.mn.y, init.pt.y),
  label = c("Weighted Mean", "Median"),
  color = c("red", "green")
)

ggplot(virginia, aes(x = long, y = lat, group = group)) +
  geom_polygon(fill = "lightblue", color = "black") +
  geom_point(
    data = t.df, aes(group = NULL, x = LONCOD, y = LATCOD)) +
  geom_point(
    data = points_to_label,
    aes(group=NULL,x = x, y = y, colour = label), size = 5) +
```

```r
geom_point(
  aes(
    x = wtd.mn.x, y = wtd.mn.y,
    colour = "Weighted Mean"),
  size = 5,
  alpha = 0.002) +
geom_point(
  aes(x = init.pt.x, y = init.pt.y, colour = "Median"), size = 5) +
scale_colour_manual(
  values = c("Weighted Mean" = "red", "Median" = "green")) +
coord_cartesian(
  xlim = c(-83.75, -75), ylim = c(36.5, 39.5)) +
theme_minimal() +
labs(
  title = "Public High Schools in Virginia",
  colour = "Calculated Point"
) +
theme(
  plot.title = element_text(hjust = 0.5),
  legend.position = c(0.9, 0.9),
  legend.background = element_rect(
    fill = "white", color = "black", size = 0.5
  ),
  legend.box.background = element_rect(
    fill = "white", color = "black", size = 0.5
  )
)
```

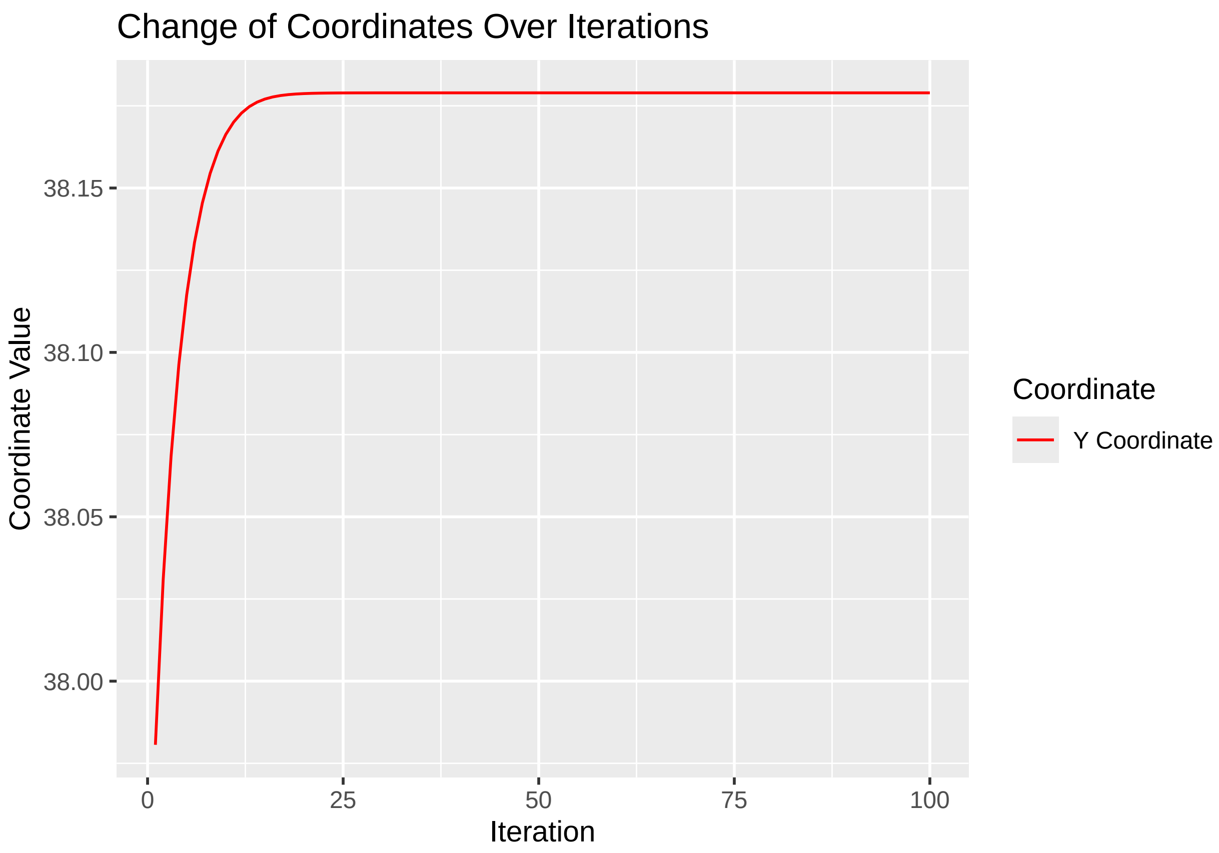


Coordinate Change over Iteration for Median Calculation

```
ggplot(coords, aes(x = iteration)) +
  geom_line(
    aes(y = x, color = "X Coordinate")) +
  labs(
    title = "Change of Coordinates Over Iterations",
    x = "Iteration",
    y = "Coordinate Value") +
  scale_color_manual(
    name = "Coordinate",
    values = c("X Coordinate" = "blue"))
```



```
ggplot(coords, aes(x = iteration)) +
  geom_line(
    aes(y = y, color = "Y Coordinate")) +
  labs(
    title = "Change of Coordinates Over Iterations",
    x = "Iteration",
    y = "Coordinate Value") +
  scale_color_manual(
    name = "Coordinate",
    values = c("Y Coordinate" = "red"))
```


Change of Coordinates Over Iterations
Coordinate Value
38.15
38.10
38.05
38.00
0
25
50
75
100
Iteration
Coordinate
Y Coordinate