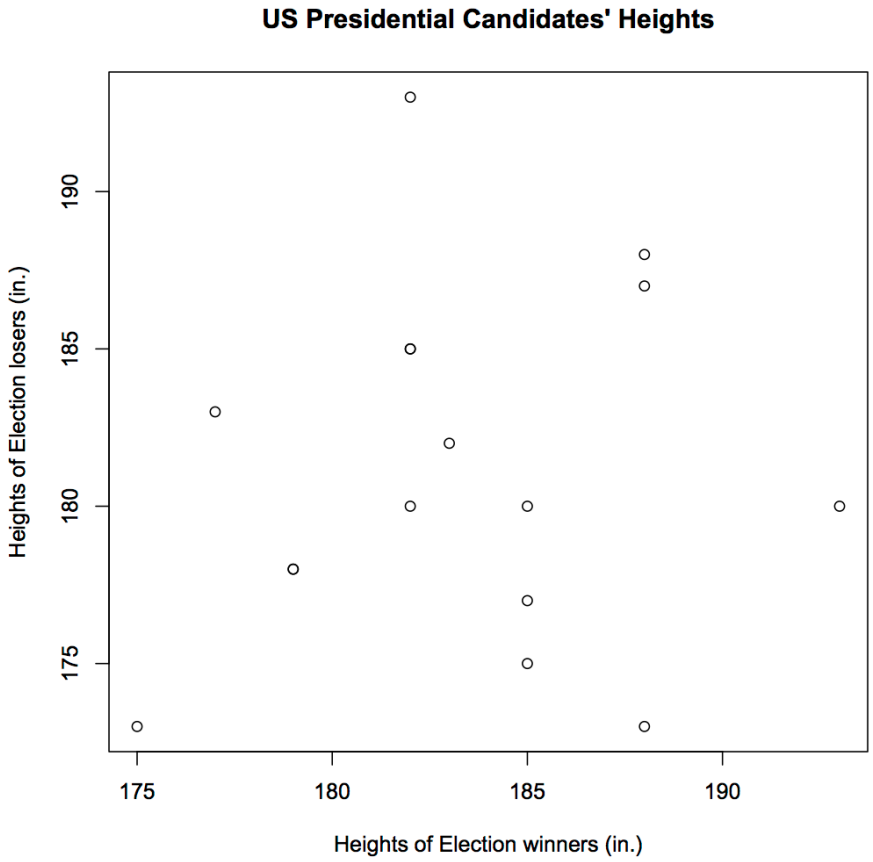
Homework 1

1.5. Find the cumulative binomial probabilities for $x = 0, 1, \ldots, 12$; $n = 12$; $p = 1/3$, using both *cumsum* and *pbinom*. What is $P(X > 7)$?

```
> # Define the problem variables
> n = 12
> p = 1/3
>
> # Calculate the binomial probabilities for x = 0, 1, ..., 12
> prob = dbinom(0:n, size=n, prob=p)
>
> # Find the cumulative values by summing the vector's values
> cumsum(prob)
 [1] 0.007707347 0.053951426 0.181122646 0.393074678 0.631520714 0.822277544
 [7] 0.933552360 0.981241568 0.996144445 0.999456196 0.999952958 0.999998118
[13] 1.000000000
>
> # Calculate the cumulative binomial probabilities directly
> pbinom(0:n, size=n, prob=p)
 [1] 0.007707347 0.053951426 0.181122646 0.393074678 0.631520714 0.822277544
 [7] 0.933552360 0.981241568 0.996144445 0.999456196 0.999952958 0.999998118
[13] 1.000000000
>
> # Find P(X > 7) knowing that P(X > 7) + P(X <= 7) == 1
> # P(X > 7) == 1 - P(X <= 7)
> 1 - pbinom(7, size=n, prob=p)
[1] 0.01875843
```

1.6. Using Wikipedia's Presidents' heights data, create a scatterplot of losers' heights vs winners' heights.

```
> # Define the problem variables
> winner.heights = c(185, 182, 182, 188, 188, 188, 185, 185, 177, 182, 182,
193, 183, 179, 179, 175)
> opponent.heights = c(175, 193, 185, 187, 188, 173, 180, 177, 183, 185,
180, 180, 182, 178, 178, 173)
>
> # Create a labeled scatterplot
> plot(winner.heights, opponent.heights, xlab="Heights of Election winners
(in.)", ylab="Heights of Election losers (in.)", main="US Presidential
Candidates' Heights")
```



1.9. Write a function $sd.n$ that will return the estimate $\widehat{\sigma}$. Try this function on the example temperature data.

```
>
# Define function sd.n
> sd.n = function(x)
+ {
+     v = var(x)
+     n = NROW(x)
+
+     sqrt(v * (n - 1) / n)
+ }
>
> # Define problem variables and test our function
> temps = c(51.9, 51.8, 51.9, 53)
> sd.n(temps)
[1] 0.4924429
```

Extra. Plot monthly average sunspot numbers against time.

```
> # Import problem data
> sunspot =
read.table("https://solarscience.msfc.nasa.gov/greenwch/spot_num.txt", header=T)
>
> # Verify it looks correct
> head(sunspot)
  YEAR MON  SSN  DEV
1 1749   1 58.0 24.1
2 1749   2 62.6 25.1
3 1749   3 70.0 26.6
4 1749   4 55.7 23.6
5 1749   5 85.0 29.4
6 1749   6 83.5 29.2
>
> # Format the time data and verify it looks correct
> Time = sunspot$YEAR + sunspot$MON / 12
> head(Time)
[1] 1749.083 1749.167 1749.250 1749.333 1749.417 1749.500
>
> # Create a linear plot of the data
> plot(Time[1:1000], sunspot$SSN[1:1000], xlab="Time (yr.)", ylab="Sunspot
Number", main="Average Sunspot Number vs. Time", type="l")
```