STATS 782 Assignment 3

Francis Tang UPI: ftan638

Due: 22 May 2019

1. An solution would be using arrows() to plot this figure.

```r
pdf("Q1.pdf", onefile=TRUE, width=7.2, height=5.5)
plot.new()
plot.window(xlim = c(-100,100), ylim = c(-100, 100))
rect(-50, -70, 50, 70,
     col = rgb(227, 100, 65, maxColorValue = 255),
     border = "white")
xl <- -30
xr <- 30
yb <- -25
yt <- 25
arrows(xl, yt, xr, yt, code = 3, angle = 125, lwd = 5,
       length = 0.4, lend = 3, col = "white")
arrows(xl, yb, xr, yb, code = 3, angle = 55,  lwd = 5,
       length = 0.4, lend = 3, col = "white")
dev.off()
```

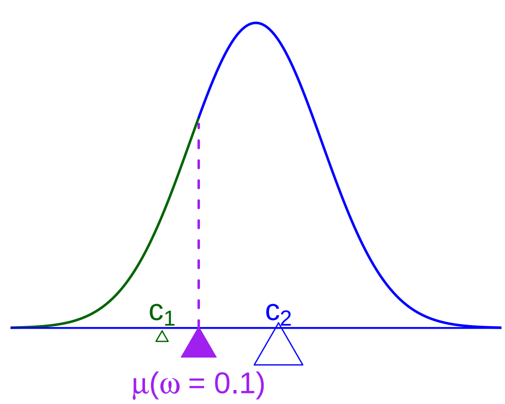
```
## pdf
##   2
```

2.

(a) Mimic the figure.

```r
pdf("Q2a.pdf", width = 7.2, height = 5)
plot.new()
plot.window(xlim = c(-3.5,3.5), ylim = c(-1,1))
x1<- seq(-5, -0.862, length=1000)
y1 <- dnorm(x1, mean=0, sd=1)
x2<- seq(-0.862, 5, length=1000)
y2 <- dnorm(x2, mean=0, sd=1)
placex <- c(-1.415,-0.862,0.342)
placey <- c(-0.013,-0.025,-0.03)
par(new = TRUE)
plot(c(x1,x2,placex,0.5), c(y1,y2, placey,-0.06),
     type = "n", xaxt='n',yaxt='n', ann=FALSE, bty ="n")
clip(-3.7, 3.7, -1, 1)
lines(x1, y1, lwd=2, col = "darkgreen")
lines(x2, y2, lwd=2, col = "blue")
abline(h = 0, lwd = 1.5, col = "blue", lend = 1)
p = dnorm(-0.862, mean = 0, sd = 1.1)
lines(rep(-0.862, 1000), seq(0, p, length = 1000),
      lty = "dashed", lwd = 2, col = "purple")
tricol <- c("darkgreen","purple","blue")
label <- c(expression("c"[1]),
           expression(paste(mu, "(", omega, " = 0.1)")),
           expression("c"[2]))
points(placex, placey, pch = c(2, 17, 2),
       xaxt = 'n', yaxt = 'n', ann = FALSE,
       col = tricol, cex = c(1, 3, 4), lwd = c(1, 1.5, 1))
mtext(label, side = 1, line = c(-3.4, -0.3, -3.4),
      at = placex, col = tricol,  cex = 1.5)
dev.off()
```

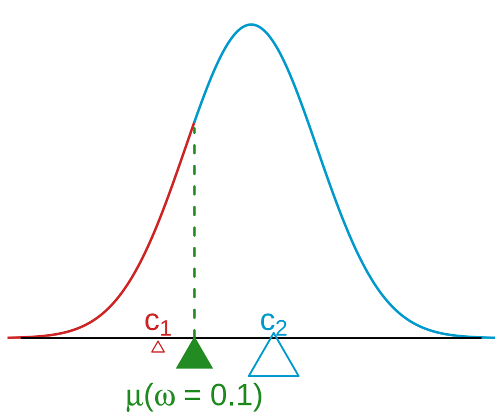
```
## pdf
##   2
```



(b) Improve the figure: the original figure above shows a high contrast in colours, for example, the blue and purple are too bright. To solve this, I chose to use some more mild colours.

```
pdf("Q2b.pdf", onefile=TRUE, width=7, height=5)
plot.new()
plot.window(xlim = c(-4, 4), ylim = c(-1,1))
x1<- seq(-5, -0.862, length=1000)
y1 <- dnorm(x1, mean=0, sd=1)
x2<- seq(-0.862, 5, length=1000)
y2 <- dnorm(x2, mean=0, sd=1)
placex <- c(-1.415,-0.862,0.342)
placey <- c(-0.013,-0.025,-0.03)
par(new = TRUE)
plot(c(x1,x2,placex,0.5), c(y1,y2, placey,-0.06),
     type = "n", xaxt='n', yaxt='n', ann=FALSE, bty ="n")
clip(-3.7, 3.7, -1, 1)
lines(x1, y1, lwd=2, col = "firebrick3")
lines(x2, y2, lwd=2, col = "deepskyblue3")
lines(c(-3.5,3.5), c(0, 0), lwd = 1.5,
      col = "black", lend = 1)
#abline(h = 0, lwd = 2, col = "dodgerblue3", lend = 1)
p = dnorm(-0.862, mean = 0, sd = 1.1)
lines(rep(-0.862, 1000), seq(0, p, length = 1000),
      lty = "dashed", lwd = 2, col = "forestgreen")
tricol <- c("firebrick3","forestgreen","deepskyblue3")
label <- c(expression("c"[1]), expression(paste(mu, "(", omega, " = 0.1)")),
           expression("c"[2]))
points(placex, placey, pch = c(2, 17, 2), xaxt = 'n', yaxt = 'n', ann = FALSE,
       col = tricol, cex = c(1, 3, 4), lwd = c(1, 1.5, 1.5))
mtext(label, side = 1, line = c(-3.4, -0.3, -3.4),
      at = placex, col = tricol, cex = 1.5)
dev.off()
```

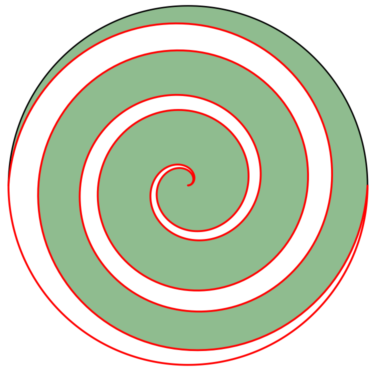
```
## pdf
##   2
```



3. Mimic the Archimedean Spiral.

```r
pdf("Q3.pdf", onefile=TRUE, width=7, height=5)
par(pty="s", bty="n", mar = c(5,5,5,5))
plot.new()
plot.window(xlim = c(-6, 6), ylim = c(-6, 6))

# draw the first Archimedean spiral
a1 <-  0.3
theta1 <- seq(0, 6 * pi, len = 400)

r1 <- a1 * theta1
x1 <- r1 * cos(theta1)
y1 <- r1 * sin(theta1)

# draw the second Archimedean spiral
a2 <- 0.36
theta2 <- seq(0, 5 * pi , len = 400)

r2 <- a2 * theta2
x2 <- r2 * cos(theta2)
y2 <- r2 * sin(theta2)

# draw the red part of the circle
theta3 <- seq(0, pi , 0.01)
r3 <- (1.8 * pi)
x3 <- r3 * cos(theta3)
y3 <- r3 * sin(theta3)

# draw the black part of the circle
theta4 <- seq(pi, 2*pi , 0.01)
r4 <- (1.8 * pi)
x4 <- r4 * cos(theta4)
y4 <- r4 * sin(theta4)

# start to fill the colour in
tx2 <- x2[length(x2) : 1]
ty2 <- y2[length(y2) : 1]
polygon(c(x1, x3, tx2), c(y1, y3, ty2),
        col = rgb(143, 188, 143, maxColorValue = 255),
        border = NA)

# finish the lines
lines(x1, y1, lwd = 1.3, col = "red")
lines(x3, y3, lwd = 1, col = "black")
lines(x2, y2, lwd = 1.3, col = "red")
lines(x4, y4, lwd = 1.3, col = "red")

dev.off()
```

```
## pdf
##   2
```



4.

(a) Check integrity.

```r
# inport dataset
datasyd <- read.csv("~/Desktop/STATS 782/datasyd.csv")
```

Now we check the annual means column by computing the means over all months.

```r
# calculate means for Jan to Dec for each year
rowMeans(datasyd[ ,4:16], na.rm = TRUE)
```

```
##   [1] 21.43077 20.29231 20.89231 21.38462 21.00000 20.60769 21.69231
##   [8] 21.25385 21.80769 21.44615 21.05385 20.70000 20.82308 20.68462
##  [15] 20.93077 21.08462 21.42308 21.29231 21.40769 21.42308 20.13846
##  [22] 20.83846 20.47692 21.16923 20.66923 21.18462 21.63077 21.28462
##  [29] 20.28462 20.97692 20.78462 20.56154 20.67692 20.36923 20.46923
##  [36] 20.91538 21.03077 20.93846 21.49231 21.33846 20.96154 20.93077
##  [43] 20.96923 20.97692 20.68462 20.93077 20.86923 21.50769 21.41538
##  [50] 21.26154 21.06923 21.37692 21.33846 21.36923 21.56154 22.03077
##  [57] 21.86154 21.10000 20.96923 21.79231 22.61538 21.68462 22.24615
##  [64] 22.53077 22.39231 21.83077 21.42308 22.50769 21.16154 22.38462
##  [71] 21.41538 21.40000 21.58462 21.51538 20.93077 21.26154 21.69231
##  [78] 21.73846 21.56923 22.09231 21.87692 22.24615 21.55385 21.73846
##  [85] 20.57692 21.63077 21.40769 21.76154 21.55385 21.27692 21.33077
##  [92] 21.40769 21.76154 21.50769 21.71538 21.66154 21.77692 21.59231
##  [99] 22.12308 22.21538 21.83077 21.62308 21.76154 21.92308 21.67692
## [106] 22.32308 22.07692 21.63077 21.72308 22.55385 22.09231 21.97692
## [113] 21.90769 21.86923 22.67692 21.67692 22.46154 22.12308 22.67692
## [120] 21.78462 22.38462 23.03846 22.13846 22.19231 22.16923 22.00000
## [127] 21.98462 22.18462 22.24615 22.84615 22.06923 22.26923 22.83846
## [134] 21.51538 22.26923 22.56154 21.82308 22.14615 22.38462 22.67692
## [141] 22.10000 22.73846 23.10000 23.05385 22.68462 23.41538 23.38462
## [148] 23.13077 22.66923 22.10000 22.93846 22.60769 22.61538 22.68462
```

```
## [155] 23.72308 23.46923 23.16923 23.83846 23.70000 23.37692 29.60000
```

```
MeanAnn <- mean(datasyd$Annual, na.rm = TRUE)
```

```
# display the annual column
datasyd$Annual
```

```
##   [1] 21.4 20.3 20.9 21.4 21.0 20.6 21.7 21.3 21.8 21.4 21.1 20.7 20.8 20.7
##  [15] 20.9 21.1 21.4 21.3 21.4 21.4 20.1 20.8 20.5 21.2 20.7 21.2 21.6 21.3
##  [29] 20.3 21.0 20.8 20.6 20.7 20.4 20.5 20.9 21.0 20.9 21.5 21.3 21.0 20.9
##  [43] 21.0 21.0 20.7 20.9 20.9 21.5 21.4 21.3 21.1 21.4 21.3 21.4 21.6 22.0
##  [57] 21.9 21.1 21.0 21.8 22.6 21.7 22.2 22.5 22.4 21.8 21.4 22.5 21.2 22.4
##  [71] 21.4 21.4 21.6 21.5 20.9 21.3 21.7 21.7 21.6 22.1 21.9 22.2 21.6 21.7
##  [85] 20.6 21.6 21.4 21.8 21.6 21.3 21.3 21.4 21.8 21.5 21.7 21.7 21.8 21.6
##  [99] 22.1 22.2 21.8 21.6 21.8 21.9 21.7 22.3 22.1 21.6 21.7 22.6 22.1 22.0
## [113] 21.9 21.9 22.7 21.7 22.5 22.1 22.7 21.8 22.4 23.0 22.1 22.2 22.2 22.0
## [127] 22.0 22.2 22.2 22.8 22.1 22.3 22.8 21.5 22.3 22.6 21.8 22.1 22.4 22.7
## [141] 22.1 22.7 23.1 23.1 22.7 23.4 23.4 23.1 22.7 22.1 22.9 22.6 22.6 22.7
## [155] 23.7 23.5 23.2 23.8 23.7 23.4   NA
```

From the result we got above, the annual means over 161 years match the annual means column, which proves the integrity of the data.

(b) Mimic the figure from 1910 to 2018.

```
# pick data from 1910 first
datasyd1910 <- datasyd[-c(1:51, 161), ]

# calculate annual mean from 1061-1990 data
datasyd1961 <- datasyd[c(103:132), ]
MeanAnn1961 <- mean(datasyd1961$Annual, na.rm = TRUE)
datasyd1910$Diff <- datasyd1910$Annual - MeanAnn1961
```

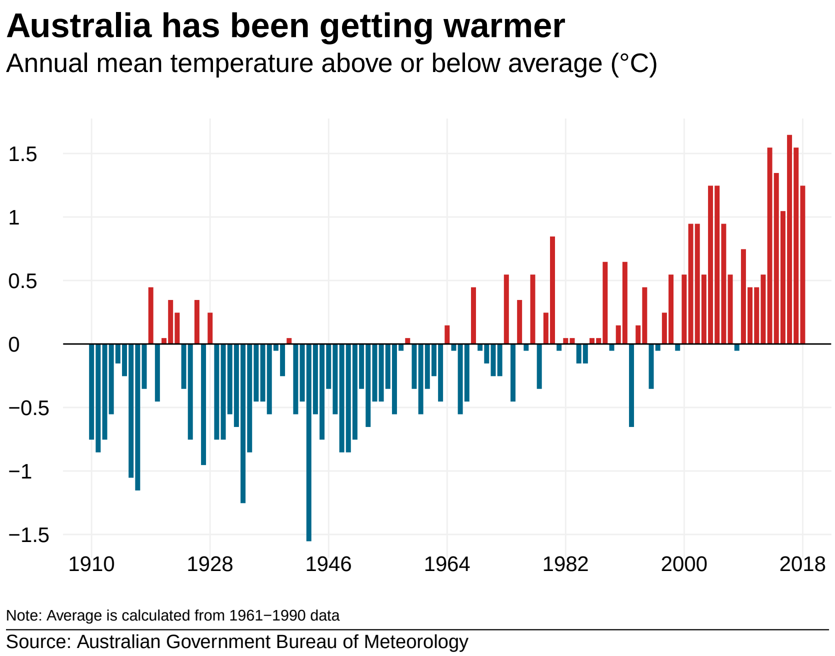
```
# prepare the whole all year's data from question (c)
datasyd1859 <- datasyd[-c(161), ]
MeanAnn1859 <- mean(datasyd1859$Annual, na.rm = TRUE)
datasyd1859$Diff <- datasyd1859$Annual - MeanAnn1961
```

The range of the temperature change is difference from the BBC figure, so I added more ticks for the y-axis.

```
#leave margin for the plot
pdf("Q4a.pdf", onefile=TRUE, width=7, height=5)
par(bty = "n", mar = c(4,5,5,2))
plot.new()
plot.window(xlim = range(datasyd1910$Year),
            ylim = range(datasyd1910$Diff))
par(new = TRUE)

# x and y axis ticks
xticks1 <- c(datasyd1910$Year[seq(1, 109, by = 18)])
yticks1 <- c(-1.5, -1, -0.5, 0, 0.5, 1, 1.5)

# draw grid in the behind
drawGrid1 <- function(){
  abline(v = xticks1, col = "gray94")
  abline(h = yticks1, col = "gray94")
}
```

```r
# plot the bars
plot(datasyd1910$Year, datasyd1910$Diff, type = "h",
     col = ifelse(datasyd1910$Diff < 0,
                  "deepskyblue4", "firebrick3"),
     lwd = 3.5, lend = 3, xaxt = 'n', yaxt = 'n',
     ylab ='', xlab ='',
     panel.first = drawGrid1())

# draw the axis and lines
axis(1, pos = -1.4, xticks1, cex.axis = 0.9,
     las = 0, font = 1, tick = FALSE)
abline(h = 0, lwd = 1, col = "black", lend = 1)
lines(c(1897, 2022), c(-2.25, -2.25), lwd = 0.8,
      col = "black", lend = 1, xpd = TRUE)

# display the text
mtext(yticks1, side = 2, at = yticks1, line = 2,
      col = 'black', cex = 0.9, adj = 0, las = 2)
mtext(c("Note: Average is calculated from 1961-1990 data"),
      side = 1,at = 1897, line = 1.5,
      col = 'black', cex = 0.7, adj = 0)
mtext(c("Source: Australian Government Bureau of Meteorology"),
      side = 1,at = 1897, line = 2.5,
      col = 'black', cex = 0.8,adj = 0)
mtext(c("Annual mean temperature above or below average (\u00B0C)"),
      side = 3,at = 1897, line = 1.5,
      col = 'black', cex = 1.2,adj = 0)
mtext(c(expression(paste(bold("Australia has been getting warmer")))),
      side = 3,at = 1897, line = 2.5,
      col = 'black', cex = 1.5,adj = 0)
dev.off()
```

```
## pdf
##   2
```



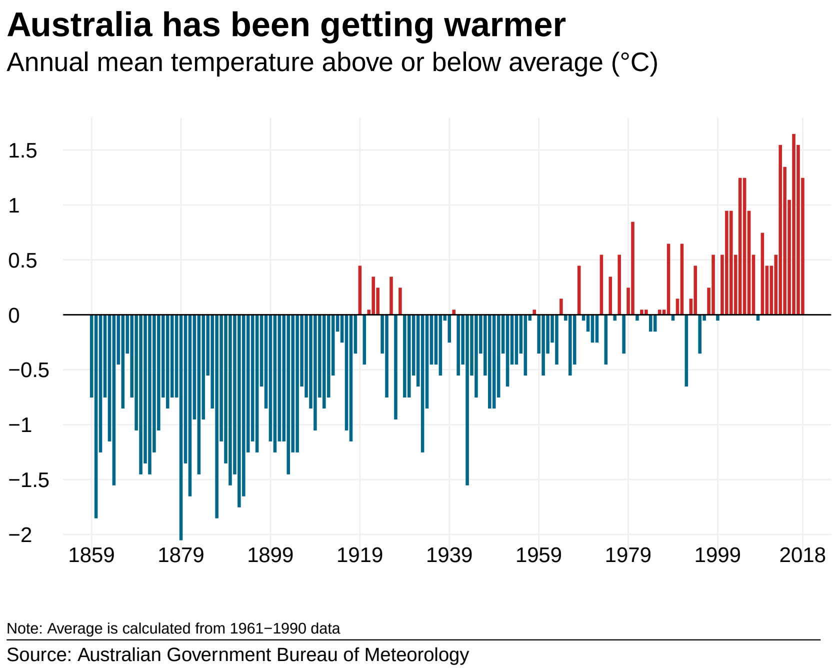
(c) Add all years data and interpret the figure.

Now we add all years data since 1859, the plot gets wider so I decreased the bar width.

```
pdf("Q4b.pdf", onefile=TRUE, width=7, height=5)
par(bty = "n", mar = c(4,5,5,2))
plot.new()
plot.window(xlim = range(datasyd1859$Year),
            ylim = range(datasyd1859$Diff))
par(new = TRUE)
xticks <- c(datasyd1859$Year[seq(1, 160, by = 20)], 2018)
yticks <- c(-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5)
drawGrid <- function(){
  abline(v = xticks, col = "gray94")
  abline(h = yticks, col = "gray94")
}
plot(datasyd1859$Year, datasyd1859$Diff, type = "h",
     col = ifelse(datasyd1859$Diff < 0,
                  "deepskyblue4", "firebrick3"),
     lwd = 2.3, lend = 3, xaxt = 'n', yaxt = 'n',
     ylab ='', xlab ='',
     panel.first = drawGrid())
axis(1, pos = -1.8, xticks,
     cex.axis = 0.9, las = 0, font = 1, tick = FALSE)
abline(h = 0, lwd = 1, col = "black", lend = 1)
lines(c(1840, 2022), c(-2.96, -2.96), lwd = 0.8,
      col = "black", lend = 1, xpd = TRUE)
mtext(yticks, side = 2, at = yticks, line = 2,
      col = 'black', cex = 0.9, adj = 0, las = 2)
mtext(c("Note: Average is calculated from 1961-1990 data"),
      side = 1,at = 1840, line = 2,
      col = 'black', cex = 0.7, adj = 0)
mtext(c("Source: Australian Government Bureau of Meteorology"),
```

```
      side = 1,at = 1840, line = 3,
      col = 'black', cex = 0.8,adj = 0)
mtext(c("Annual mean temperature above or below average (\u00B0C)"),
      side = 3,at = 1840, line = 1.5,
      col = 'black', cex = 1.2,adj = 0)
mtext(c(expression(paste(bold("Australia has been getting warmer")))),
      side = 3,at = 1840, line = 2.5,
      col = 'black', cex = 1.5,adj = 0)
dev.off()
```

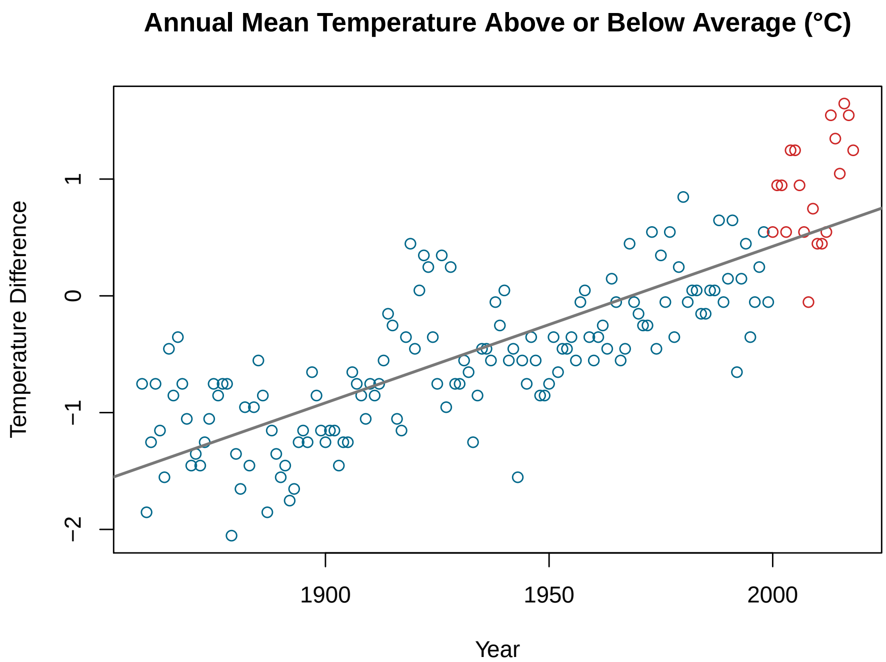
```
## pdf
##   2
```



It is not sufficient enough to see the trend and information behind the figure, so I fit a simple linear regression based on the data and the scatter plot is displayed below with the regression line.

```
pdf("Q4c.pdf", onefile=TRUE, width=7, height=5)
par(mar = c(4,5,4,2))
plot(datasyd1859$Year, datasyd1859$Diff, pch = 21,
     col = ifelse(datasyd1859$Year < 2000, "deepskyblue4", "firebrick3"),
     xlab = "Year", ylab = "Temperature Difference")
title("Annual Mean Temperature Above or Below Average (\u00B0C)", line = 2)
abline(lm(formula = Diff ~ Year, data = datasyd1859), lwd = 2, col = "gray47")
dev.off()
```

```
## pdf
##   2
```



From the two plots we generated above, we can conclude the following statements:

1. The temperature of Australia keeps arising during the last 150 years.
2. Australia gets warm more rapidly since 2000, which has been shown on the scatter plot, the slope of temperature increase gets larger and the scatters are even above the regression line.
3. Generally speaking, the data from 1859 til now can roughly fit a simple linear regression.

(d) Predict temperature in 2030 and give 95% prediction interval.

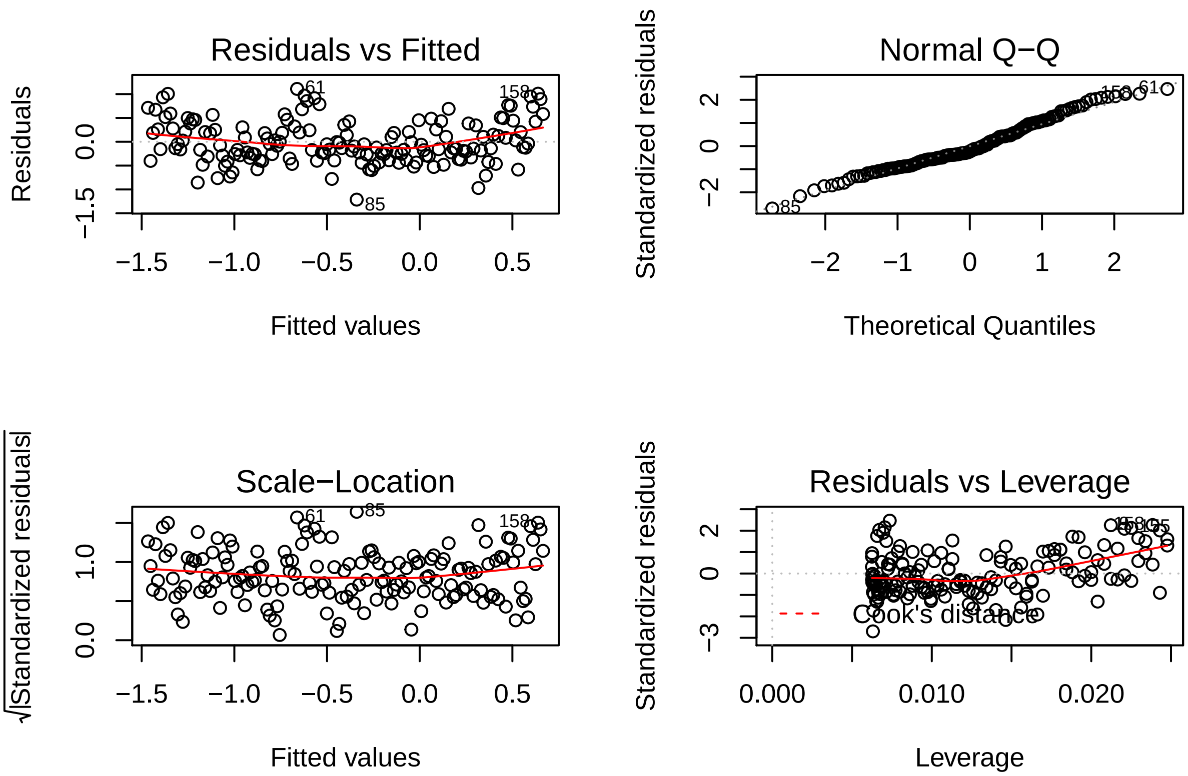
```
# first we fit a simple linear regression model
TempDiff.lm = lm(formula = Diff ~ Year, data = datasyd1859)
# check the summary data
summary(TempDiff.lm)
```

```
##
## Call:
## lm(formula = Diff ~ Year, data = datasyd1859)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2134 -0.2928 -0.1057  0.2845  1.1082
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.638e+01  1.496e+00  -17.63   <2e-16 ***
## Year         1.340e-02  7.717e-04   17.36   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4509 on 158 degrees of freedom
## Multiple R-squared:  0.6562, Adjusted R-squared:  0.654
## F-statistic: 301.5 on 1 and 158 DF,  p-value: < 2.2e-16
```

```r
# use ANOVA
anova(TempDiff.lm)
```

```
## Analysis of Variance Table
##
## Response: Diff
##            Df Sum Sq Mean Sq F value    Pr(>F)
## Year        1 61.300  61.300  301.55 < 2.2e-16 ***
## Residuals 158 32.119   0.203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# use diagnostic plots
par(mfrow=c(2,2))
plot(TempDiff.lm)
```



After fit a linear regression model on the data, we can see that "Year" regressor shows as a very significant regressor to the model and the diagnostic plots prove it as well. So there exist a significant linear regression between year and temperature change.

The next step is to use the model to predict temperature in 2030. First we are going to predict the time difference then add it into the mean of 1961-1990 data.

```r
# use the linear model to make prediction for 2030
pred2030 <- predict(TempDiff.lm, newdata = data.frame(Year = 2030),
        interval="prediction", level = 0.95)
temp2030 = MeanAnn1961 + pred2030
temp2030
```

```
##        fit      lwr      upr
## 1 22.97935 22.07524 23.88346
```

The data above shows us the prediction of 2030 temperature is 22.98 degrees in Celcius. And the 95% prediction interval is between (22.075, 23.883).

(e) Bonus

Here we calculate the probablity of greater than 6.5 degrees in the future:

```
1 - pexp(6.5, rate = (1/2.17), lower.tail = TRUE, log.p = FALSE)
```

```
## [1] 0.05001703
```

So we can see that the temperature is 2.17 at the year, substitute 2.17 into the linear model we fit before:

```
yearPredict <- (2.17 + 2.638e+01) / 1.340e-02
yearPredict - 2100
```

```
## [1] 30.59701
```

The answer is 30.6 years.