

Regression and Residuals Visualization

Visualization Code

This is a visualization of how regression lines and residuals are related. “Looseness” and “Tightness” of a string or rubber band is the analogy being used with the residuals (i.e. red residual lines are the most “stretched” or “tight” lines as they are furthest from the regression line and are “pulling” the most). It is useful to think that each point has equal length of string or rubber band.

This is mainly to help students visualize how the regression line represents a balance or equilibrium of the residuals for all of the points with respect to the regression line (i.e. compare the color balance on either side of the regression line). However, these plots are also useful for seeing how different outliers might influence the regression line. For instance, some outliers are outliers w.r.t x values but may be very close to the regression line, while others may be outliers w.r.t to y values and may be very far from the regression line. Another thing to notice is how the sample size relates to the influence of an outlier (i.e. outliers have more influence on the regression line with smaller sample sizes).

```
colors = c("green", "blue", "purple", "red")
xpt = c(140, 190)
ypt = c(150, 240)
n = c(5, 25, 100)

for (k in 1:length(n)){
  df = rdist(r=0.8, n=n[k], xp=c(190, 10), yp=c(190, 10), rs=1)
  m_b = get_stat(df)
  rplot(df)
  rrplot(df, colors)
  for(i in 1:length(xpt)){
    for(j in 1:length(ypt)){
      rrplot2(df, colors, m_b, pt = c(xpt[i], ypt[j]))
    }
  }
}
```

















