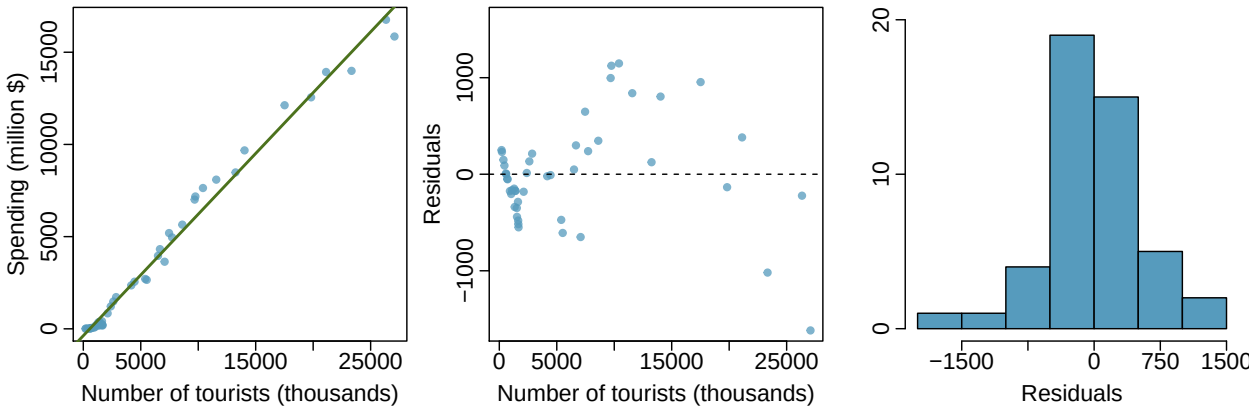
Examples: Simple Linear Regression

Math 107, Spring 2016

Tourism spending

The Association of Turkish Travel Agencies reports the number of foreign tourists visiting Turkey and tourist spending by year. Three plots are provided: a scatterplot showing the relationship between these two variables along with the least squares fit, a residual plot, and a histogram of residuals.



1. Describe the relationship between number of tourists and spending.

2. What are the explanatory and response variables?

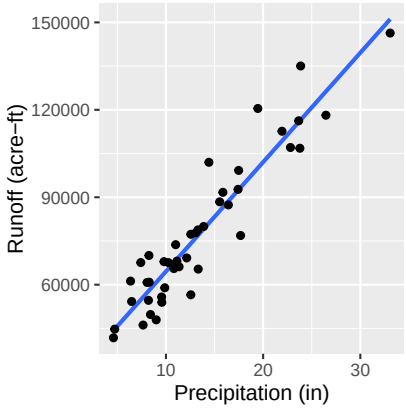
3. Why might we want to fit a regression line to these data?

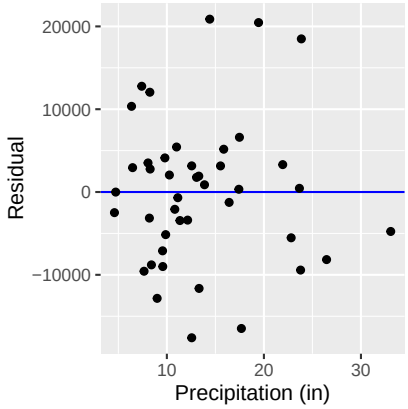
4. Do the data meet the conditions required for fitting a least squares line? In addition to the scatterplot, use the residual plot and histogram to answer this question.

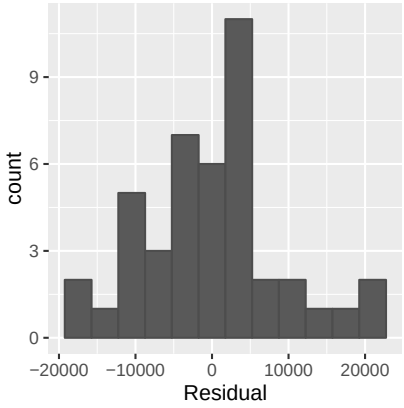
Stream runoff in California

In recent years, California has experienced severe to exceptional drought conditions. As of May 26, the U.S. Drought Monitor (http://droughtmonitor.unl.edu/) shows that 61% of the state is currently experiencing severe to exceptional drought. One factor affecting water availability in Southern California is stream runoff from snowfall. If runoff could be predicted, engineers, planners, and policy makers could do their jobs more effectively because they would have an estimate as to how much water is entering the area.

The data set `water.txt`[1] (available on the course webpage) compares the stream runoff (in acre-feet) of a river near Bishop, California (due east of San Jose) with snowfall (in inches) at a site in the Sierra Nevada mountains. Regression results and associated plots for a simple linear regression model fit to these data are shown below. Use this output to answer the following questions.








```
model <- lm(Runoff ~ Precip, data = water)
summary(model)
```

```
##
## Call:
## lm(formula = Runoff ~ Precip, data = water)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17603.8  -5338.0    332.1   3410.6  20875.6
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27014.6     3218.9   8.393 1.93e-10 ***
## Precip        3752.5      215.7  17.394  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8922 on 41 degrees of freedom
## Multiple R-squared:  0.8807, Adjusted R-squared:  0.8778
## F-statistic: 302.6 on 1 and 41 DF,  p-value: < 2.2e-16
```

1. Describe the apparent relationship between stream runoff and snowfall based on the scatterplot.

[1]Source: Weisberg, S (2014). *Applied Linear Regression.* Wiley.

2. Report the equation for the regression line fit using R.

3. Interpret the estimated slope in the context of the problem.

4. Interpret the estimated y-intercept in the context of the problem.

5. In the winter of 2013-2014, the site only received 4.5 inches of snowfall. What do you predict will be the associated runoff? Do you have any hesitations about making this prediction?

6. Report the value of R^2 and interpret this in the context of the problem.

7. Is snowfall a significant predictor of stream runoff? Carry out the appropriate hypothesis test, being sure to complete all five steps.

8. Calculate a 95% confidence interval for the slope. Does this confidence interval agree with your hypothesis test?