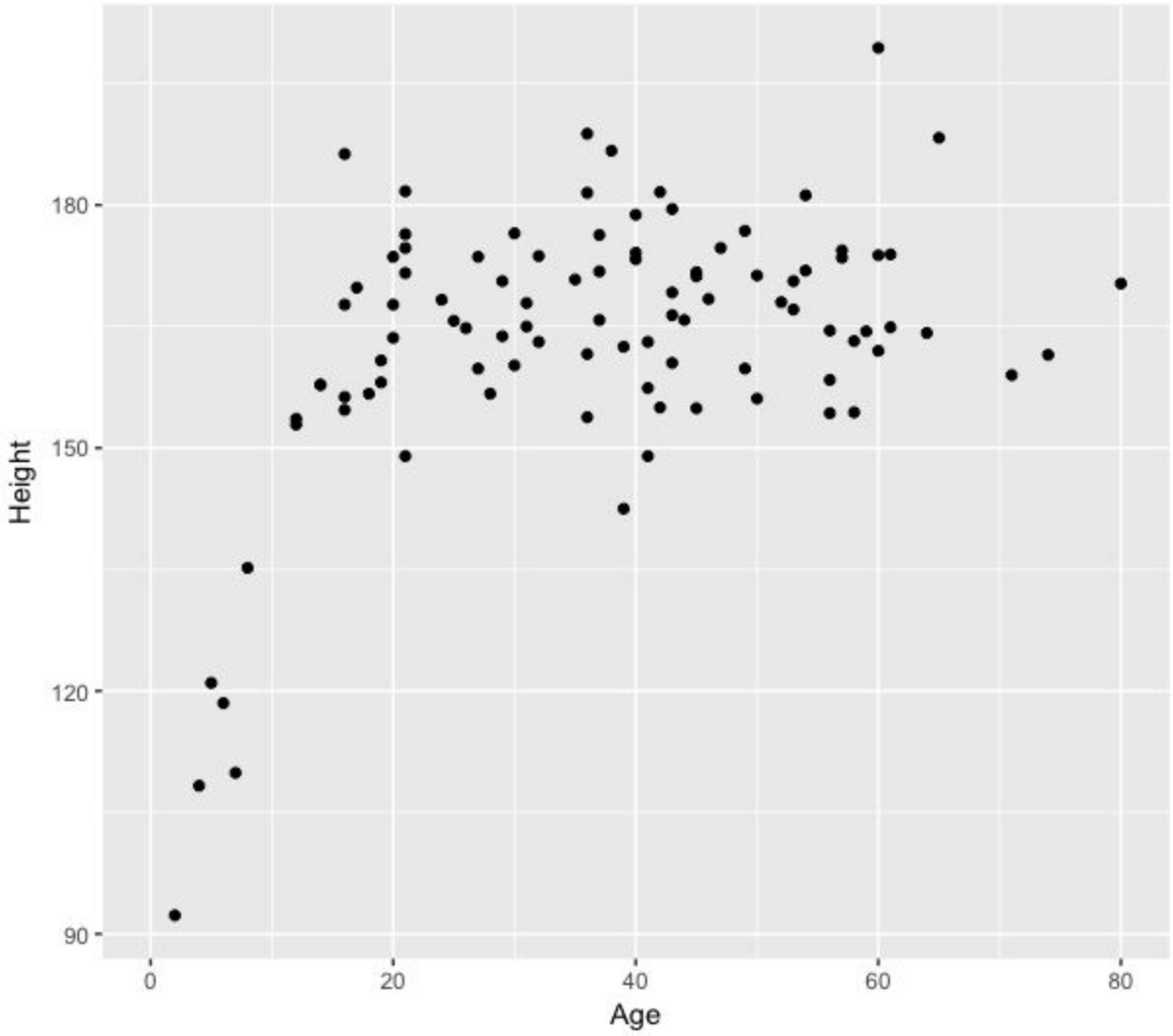
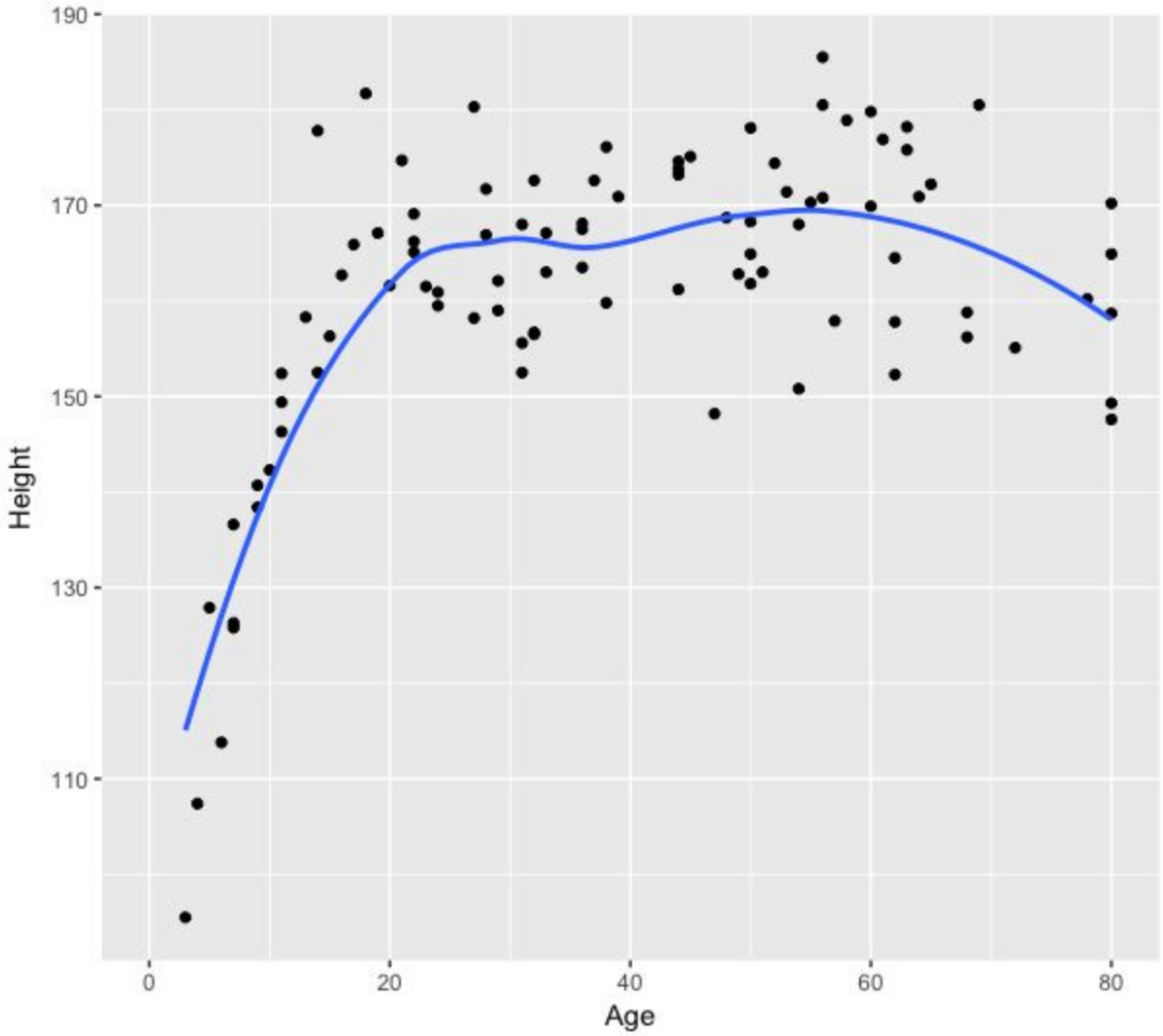
Regression and confidence

1. What is a function? Trace trajectory of height vs age.
2. Hand out the sheet protectors. Hand out data plots of the variables in the function: height and age.
    a. Take a sample of about 100 cases from NHANES and make a data plot.
    b. library(NHANES)
    c. gf_point(Height ~ Age, data = sample(NHANES, 100))
    d. ;



    Print out a dozen or so plots, each of a different sample of 100 from NHANES. Each group inserts the page into their paper protector, then draws the function with a sharpie. Ask the students to mark down a few points showing typical height at different ages, then connect them with a smooth curve.
3. Collect the paper protectors. Stack them on top of one another to show the variation in the student functions.

4. On the computer, make a plot of a sample and a smooth curve constructed by the computer.
    a. `gf_point(Height ~ Age, data = sample(NHANES, 100)) %>% gf_smooth(se = FALSE)`



    Maybe do this several times to show how the function jumps around, just like the students' plots.
    b. Discuss the range of reasonable functions
    c. Then add in se = TRUE to see the confidence band

5. Other functions. Use to discuss whether they are better or worse.
    a. `gf_point(Height ~ Age, data = sample(NHANES, 100)) %>% gf_abline(intercept=175, slope = - 2)`
    b. `gf_point(Height ~ Age, data = sample(NHANES, 100)) %>% gf_lm(color = ~ Age > 20)`

```
c. gf_point(Height ~ Age, data = sample(NHANES, 100)) %>%
   gf_smooth(color = ~ Age > 20, group = ~ Gender)
```

P-value example

1. Pull out only the people over 20 years old. Select a small sample, say n = 5
2. Fit a linear trend, to data with Age shuffled. Plot out the confidence interval around the linear fit with the randomized data.
3. Plot out the linear trend for the actual data.
4. See whether the actual-data modelfits in the confidence bands for the null-hypothesis model..
5. Increase the sample size until it doesn't
6. Explain that the p-value quantifies how different the actual data is from the shuffled data.