# CORRELATION HOMEWORK

TEXT: 12.1, 12.2, 12.3,

| LAST NAME | FIRST NAME | DATE |
|---|---|---|
| | | |

Consider the datasets `state.center$x` and `state.center$y` with longitudes and lattitudes of US states respectively, and `state.area` with areas of US states in square miles.

**1.** Find the linear correlation coefficient for datasets `state.center$x` and `state.center$y`. Describe the sign and the strength of the linear correlation in English.

**2.** Find the linear correlation coefficient for `state.center$x` and `state.area`. What can be said about the areas of Eastern states (the ones with higher $x$ coordinates)?

**3.** Find an equation in slope-intercept form for the line of best fit predicting `state.area` given `state.center$x`

**4.** Use your equation to predict the area of a state with longitude $x = -100$

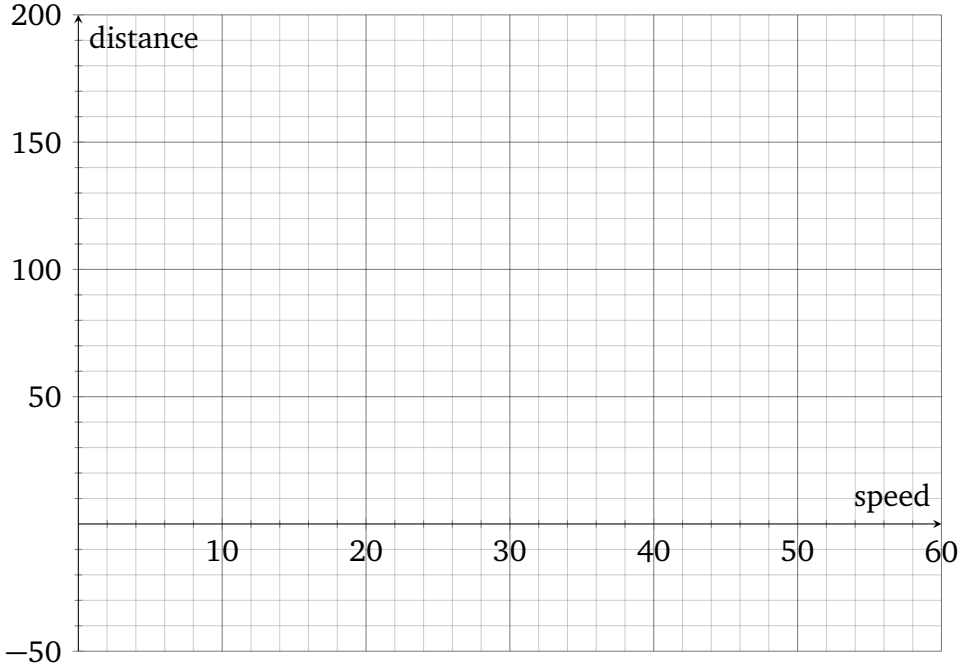
Consider the datasets `cars$speed` and `cars$dist`, with speeds in mph and corresponding stopping distances in feet.

**5.** Find the linear correlation coefficient for datasets `cars$speed` and `cars$dist`. Do they seem strongly correlated? What can be said about the stopping distances of cars which go faster?

**6.** Find an equation in slope-intercept form for the line of best fit predicting the stopping distance given the speed of a car.

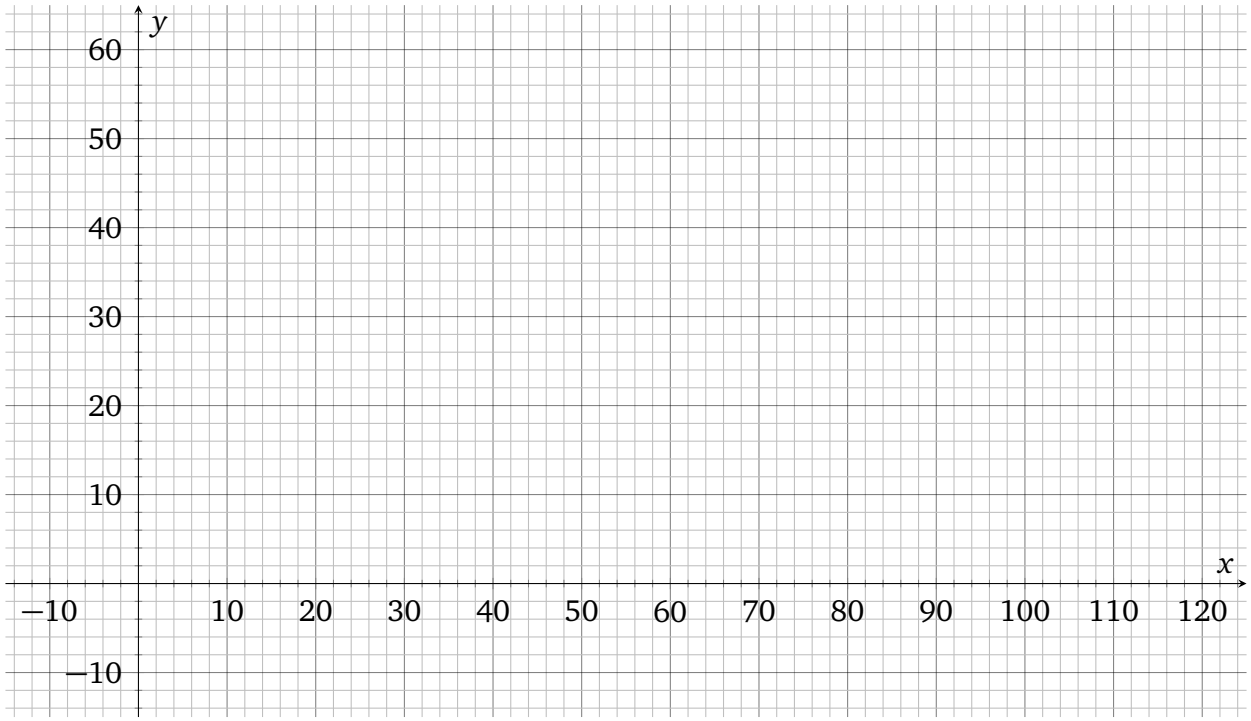
**7.** Use your equation to predict the stopping time of a car that moves at 5 mph, and the stopping time of a car that moves at 50 mph.

**8.** Plot the line of best fit on the grid below. Label the following points on the line: the $y$-intercept, the point corresponding to 5 mph, and the point corresponding to 50 mph.



Consider the following data, where $x$ represents the year since 1900 and $y$ represents the population of a nation state in millions of people. For example, the point $(20, 3.8)$ corresponds to the population of 3.8 million in the year 1920, while the point $(100, 34)$ records the population of 34 million in year 2000.

| $x$ | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| $y$ | 34 | 15 | 42 | 40 | 48 |




**9.** Plot the points on the grid.

**10.** Without computing anything, **use a straightedge** to draw an **approximate** line of best fit. This line should be straight, should extend through the given coordinate grid, and be as close to as many points as possible. It does not have to pass through any one data point, though it might just by chance.

**11.** Use your graph (**not** R) to find an equation of your line in the slope-intercept form.

**12.** Use your line to predict the population in year 2030.

**13.** Use your line to estimate the sum of squares of residuals.

# ANSWERS

**1**. $r = -0.07900579$, very weak negative correlation.

**2**. $r = -0.5753457$, negative association means that states further to the East tend to have smaller areas.

**3**. $\hat{a} = -3139x - 217871$

**5**. $r = 0.8068949$, large positive correlation means that cars with higher speed will tend to have larger stopping distances.

**7**. 2.081 feet for 5 mph and 179.021 feet for 50 mph