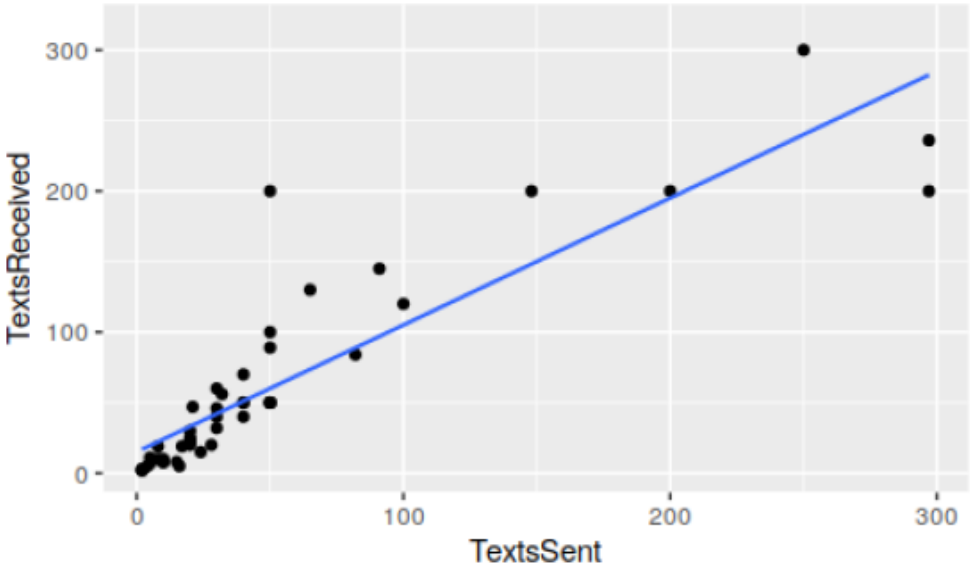
Regression Computer Lab

Max Kurzweil

2/7/2025

```
library(readr)
library(mosaic)
library(ggformula)

Spring25Data <- read_csv("Spring25Data.csv")
## Rows: 50 Columns: 18
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr (10): ToppingPref, GiveUp, TextOften, Active, Generous, Gender, CoffeeFreq,
StudyLocationOriginal, Personality, StudyLocation
## dbl (8): SpendOnYou, LongestRun, PizzaToppings, TextsSent, TextsReceived,
GroupText, SpeedAlone, SpeedPassenger
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

head(Spring25Data)

## # A tibble: 6 × 18
##   SpendOnYou LongestRun PizzaToppings ToppingPref  GiveUp  TextsSent
TextsReceived
##        <dbl>      <dbl>         <dbl> <chr>        <chr>       <dbl>      <dbl>
## 1        900         NA             0 Only cheese  Meat           NA         NA
## 2       1000          2             4 A bit of both Meat          82         84
## 3        500          5             3 Only cheese  Dairy          10          8
## 4        500          1             2 A bit of both Dessert       50        200
## 5       1000          1             4 A bit of both Caffeine      50        100
## 6       1000         10             4 Mostly meat  Dessert        50         50
## # ℹ 11 more variables: GroupText <dbl>, TextOften <chr>, Active <chr>,
## #   Generous <chr>, Gender <chr>, SpeedAlone <dbl>, SpeedPassenger <dbl>,
## #   CoffeeFreq <chr>, StudyLocationOriginal <chr>, Personality <chr>,
## #   StudyLocation <chr>

gf_point(TextsReceived ~ TextsSent, data =Spring25Data) %>% gf_lm()

## Warning: Removed 1 rows containing non-finite values (stat_lm).

## Warning: Removed 1 rows containing non-finite values (geom_point).
```



According to the scatterplot, most students sent between 0 and 100 text messages, with a concentration in the 0-50 range. There appears to be a strong positive linear relationship between texts sent and texts received.

```
cor(TextsReceived ~ TextsSent, data =na.omit(Spring25Data))

## [1] 0.8885726
```

The correlation between texts sent and texts received is $r = 0.889$, indicating a strong positive linear relationship (values above 0.8 are generally considered strong).

```
reg <- lm(TextsReceived ~ TextsSent, data =Spring25Data)
summary(reg)

## lm(formula = TextsReceived ~ TextsSent, data = Spring25Data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##  -82.368  -13.906  -10.100    4.919  139.900
##
## Coefficients:
##             Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept) 15.10665    5.75781    2.624    0.0117 *
## TextsSent    0.89987    0.06679   13.473   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.42 on 47 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.7943,  Adjusted R-squared:  0.7901
## F-statistic: 181.5 on 1 and 47 DF,  p-value: < 2.2e-16
```

The slope of the regression model is $b1=0.900$. For every additional text message sent, we expect to receive approximately 0.9 additional text messages. The intercept is $b0=15.107$, which means that when zero text messages are sent, we expect to receive approximately 15 text messages (though this interpretation may not be practically meaningful if zero is outside the observed data range).

Prediction for 400 texts sent:

```
15.10665 + 0.89987 * 400
```

```
## [1] 375.0546
```

The predicted number of texts received for someone who sends 400 texts is approximately 375.

Caution about extrapolation: People should be hesitant about this prediction because 400 texts sent is far outside the range of our observed data (which appears to be roughly 0-200). This is extrapolation, which can be unreliable. I would expect the residual for this prediction to be potentially large, though the direction (positive or negative) is uncertain.

The model accounts for approximately 79.4% of the variability in the number of texts received ($R^2=0.7943$), indicating that texts sent is a strong predictor of texts received.