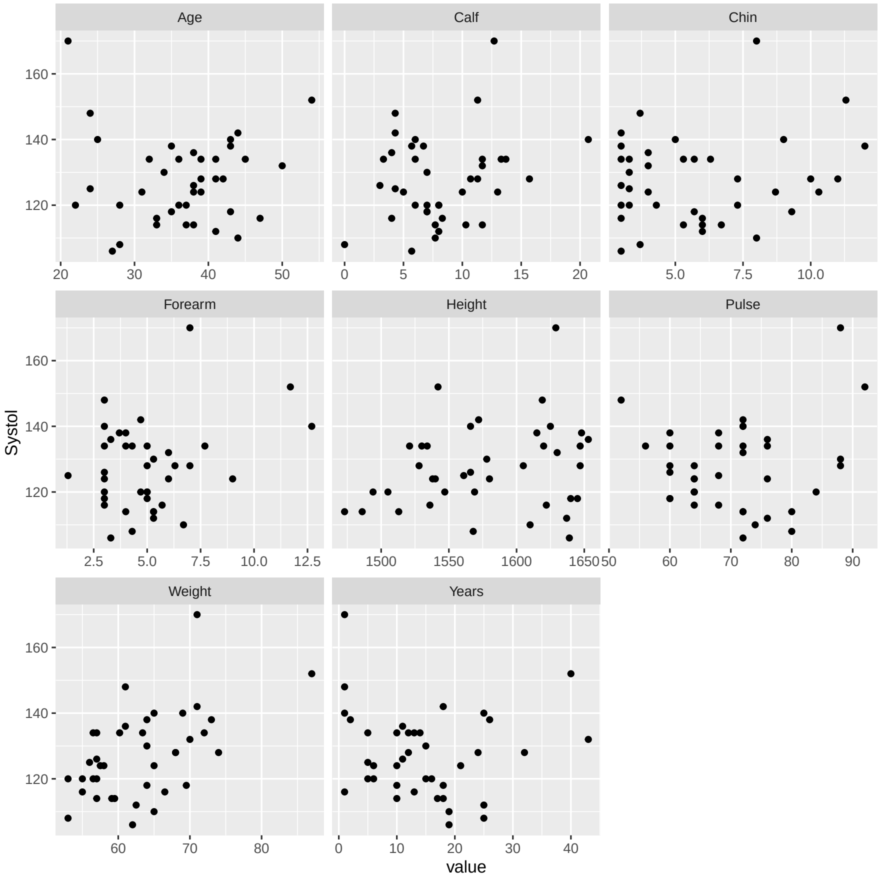
Assignment 5: Under (blood) pressure

Minh Tran

2021-05-28

Exercise 1

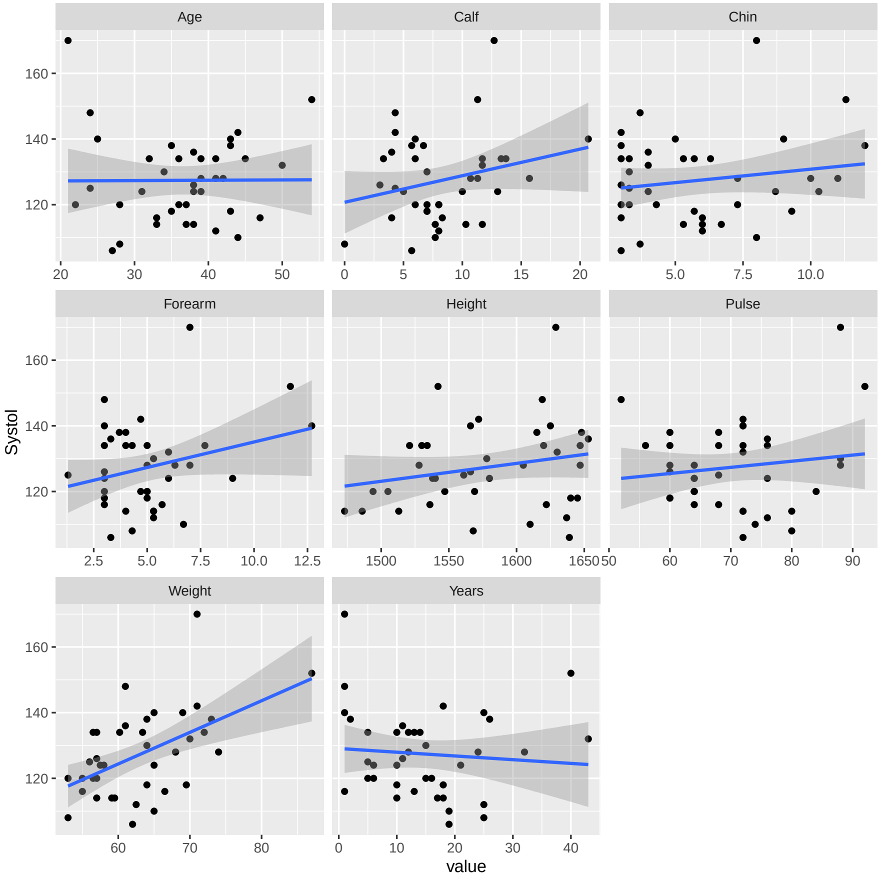
```
blood_pressure %>%
  gather(Age:Pulse, key = "measurement", value = "value") %>%
  ggplot() +
    geom_point(mapping = aes(x = value, y = Systol)) +
    facet_wrap(~ measurement, scales = "free_x")
```



Exercise 2

```
blood_pressure %>%
  gather(Age:Pulse, key = "measurement", value = "value") %>%
  ggplot() +
  geom_point(mapping = aes(x = value, y = Systol)) +
  facet_wrap(~ measurement, scales = "free_x") +
  geom_smooth(mapping = aes(x = value, y = Systol), method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



i. ***It looks like there is a negative correlation between Years and Systol.***

ii. ***The variables that show a moderate to strong positive correlation with systol include Calf, Forearm, and Weight.***

Exercise 3

```
blood_pressure_updated <- blood_pressure %>%
  mutate(urban_frac_life = Years / Age)
```

Exercise 4

```
systol_urban_frac_model <- lm(Systol ~ urban_frac_life, data = blood_pressure_updated)
```

Exercise 5

```
systol_urban_frac_model %>%
  tidy()
```

term	estimate	std.error	statistic	p.value
(Intercept)	133.49572	4.038011	33.059770	0.0000000
urban_frac_life	-15.75182	9.012962	-1.747686	0.0888139

```
systol_urban_frac_model %>%
  glance() %>%
  glimpse()
```

```
## Rows: 1
## Columns: 12
## $ r.squared     <dbl> 0.07625642
## $ adj.r.squared <dbl> 0.05129038
## $ sigma         <dbl> 12.76966
## $ statistic     <dbl> 3.054406
## $ p.value       <dbl> 0.08881392
## $ df            <dbl> 1
## $ logLik        <dbl> -153.6478
## $ AIC           <dbl> 313.2957
## $ BIC           <dbl> 318.2864
## $ deviance      <dbl> 6033.372
## $ df.residual   <int> 37
## $ nobs          <int> 39
```

Exercise 6

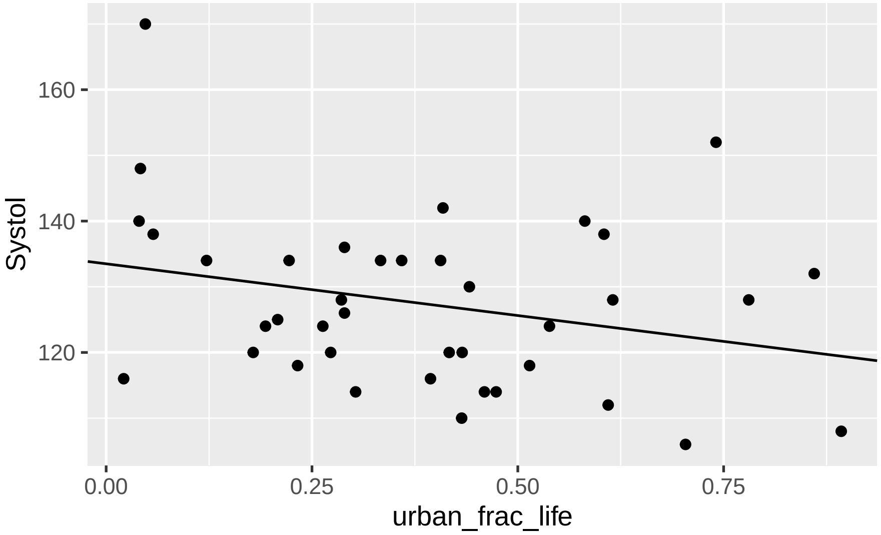
```
systol_urban_frac_df <- blood_pressure_updated %>%
  add_predictions(systol_urban_frac_model) %>%
  add_residuals(systol_urban_frac_model)
```

i. ***The column that has the predictions made by the model is 'pred'.***

ii. ***The column that has the residuals for each observation os 'resid'.***

Exercise 7

```
ggplot(systol_urban_frac_df) +
  geom_point(mapping = aes(x = urban_frac_life, y = Systol)) +
  geom_abline(slope = systol_urban_frac_model$coefficients[2], intercept = systol_urban_frac_mo
```



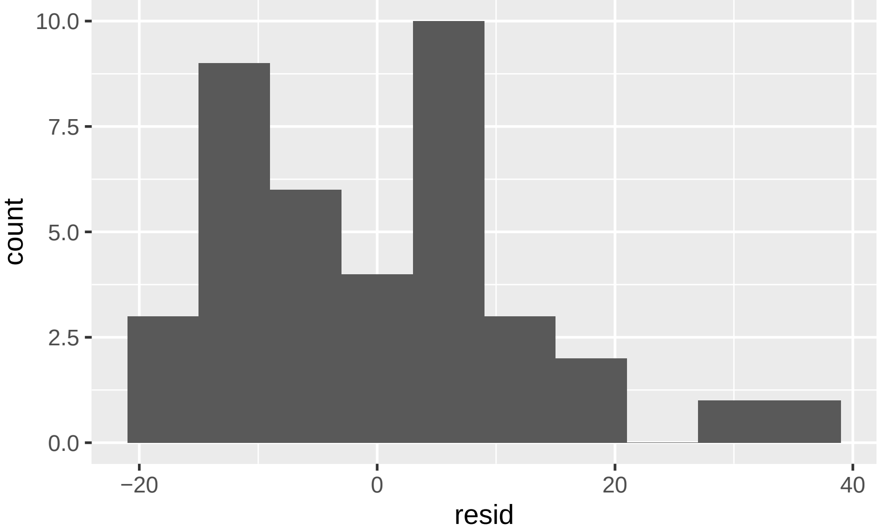
It looks like the model meets the linearity condition.

Exercise 8

i. It looks like there is constant variability in terms of the distribution of points along the regression line. As a result, the constant variability condition is met.

Exercise 9

```
ggplot(systol_urban_frac_df) +
  geom_histogram(aes(x = resid), binwidth = 6)
```



i.The distribution of the residuals are not normally distributed. It looks like it is right-skewed.

ii. The histogram suggests that the condition of normal residuals is not met since the residuals are really skewed.

Exercise 10

```r
systol_weight_model <- lm(Systol ~ Weight, data = blood_pressure_updated)
```

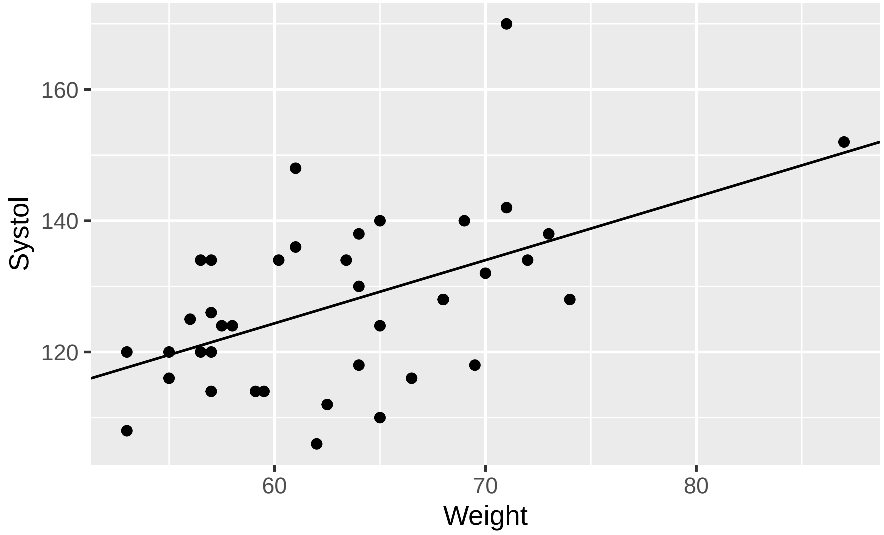
```r
systol_weight_model %>%
  glance() %>%
  glimpse()
```

```
## Rows: 1
## Columns: 12
## $ r.squared     <dbl> 0.2718207
## $ adj.r.squared <dbl> 0.2521402
## $ sigma         <dbl> 11.33764
## $ statistic     <dbl> 13.81166
## $ p.value       <dbl> 0.0006654447
## $ df            <dbl> 1
## $ logLik        <dbl> -149.009
## $ AIC           <dbl> 304.0181
## $ BIC           <dbl> 309.0088
## $ deviance      <dbl> 4756.056
## $ df.residual   <int> 37
## $ nobs          <int> 39
```

The R-squared value in the systol weight model is higher (0.27 vs 0.08) which indicates that it is the better model in terms of predicting systol .

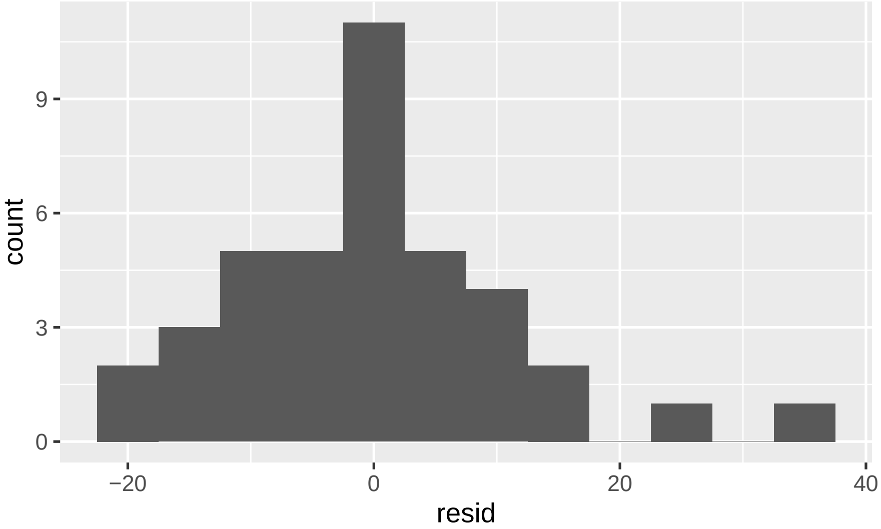
Exercise 11

```r
ggplot(systol_weight_model) +
  geom_point(mapping = aes(x = Weight, y = Systol)) +
  geom_abline(slope = systol_weight_model$coefficients[2], intercept = systol_weight_model$coe
```



```r
systol_weight_df <- blood_pressure_updated %>%
  add_predictions(systol_weight_model) %>%
  add_residuals(systol_weight_model)
```

```r
ggplot(systol_weight_df) +
  geom_histogram(aes(x = resid),binwidth = 5)
```



It looks like the there is a linear relationship between systol and weight, so the linearity condition is met. Also, there is constant variability of the distribution of data points along the regression line, so the constant variability condition is met. Even though the normality of the residuals is not perfect, it is close enough of showing a normal distribution by which the condition of normal residuals are met.

Exercise 12

Comparing the r-squared values, the systol_weight model is better at predicting systol since it has a higher r-squared value. Both models met the linearity condition as well as the constant variability condition. However, the systol_weight model met the normality of residuals condition better than the systol_urban model which almost violated that condition since the residuals are not normal at all.