Assignment 5: Under (blood) pressure

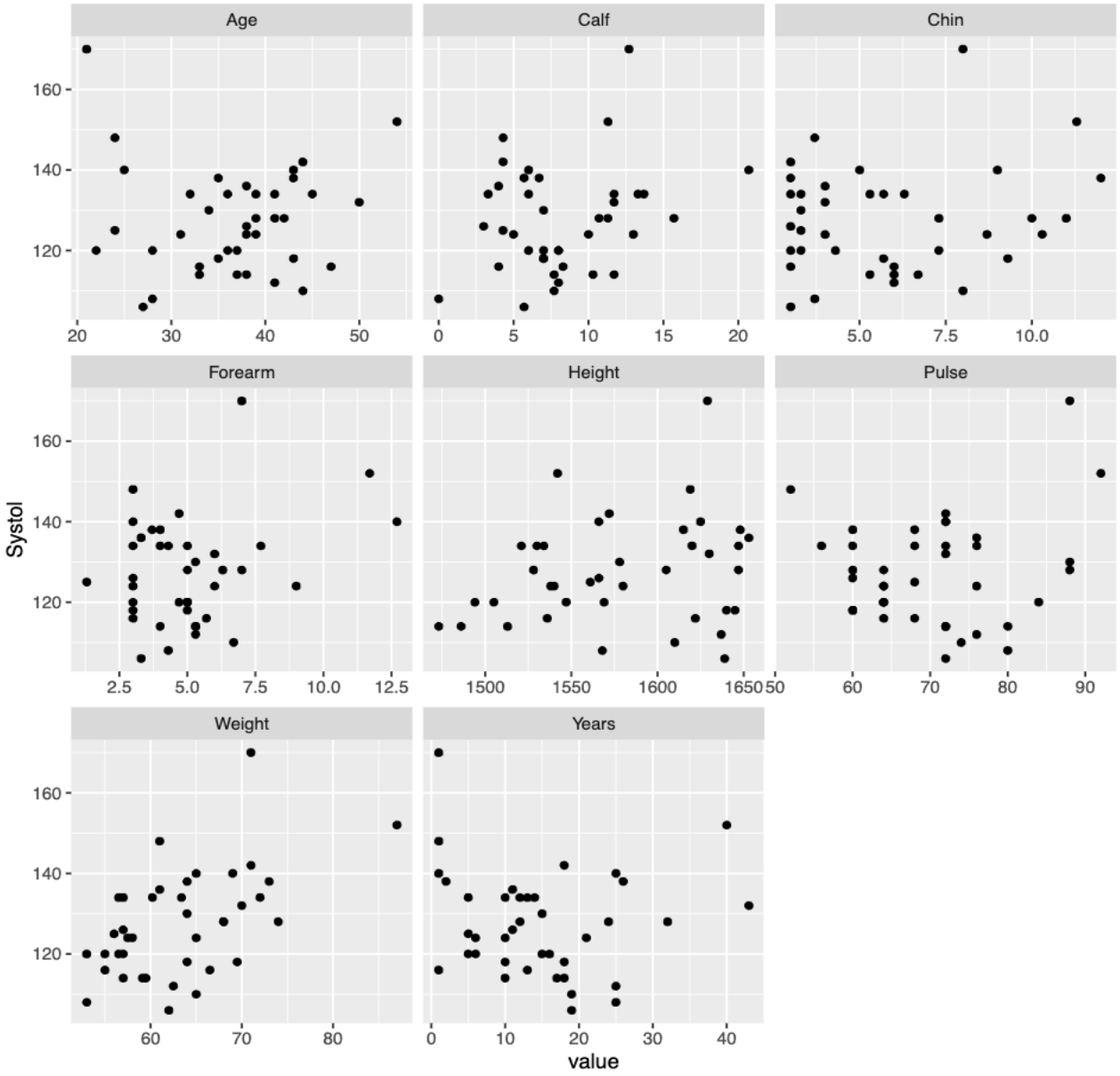
Alizeh Murtaza

2023-02-24

Exercise 1

i.

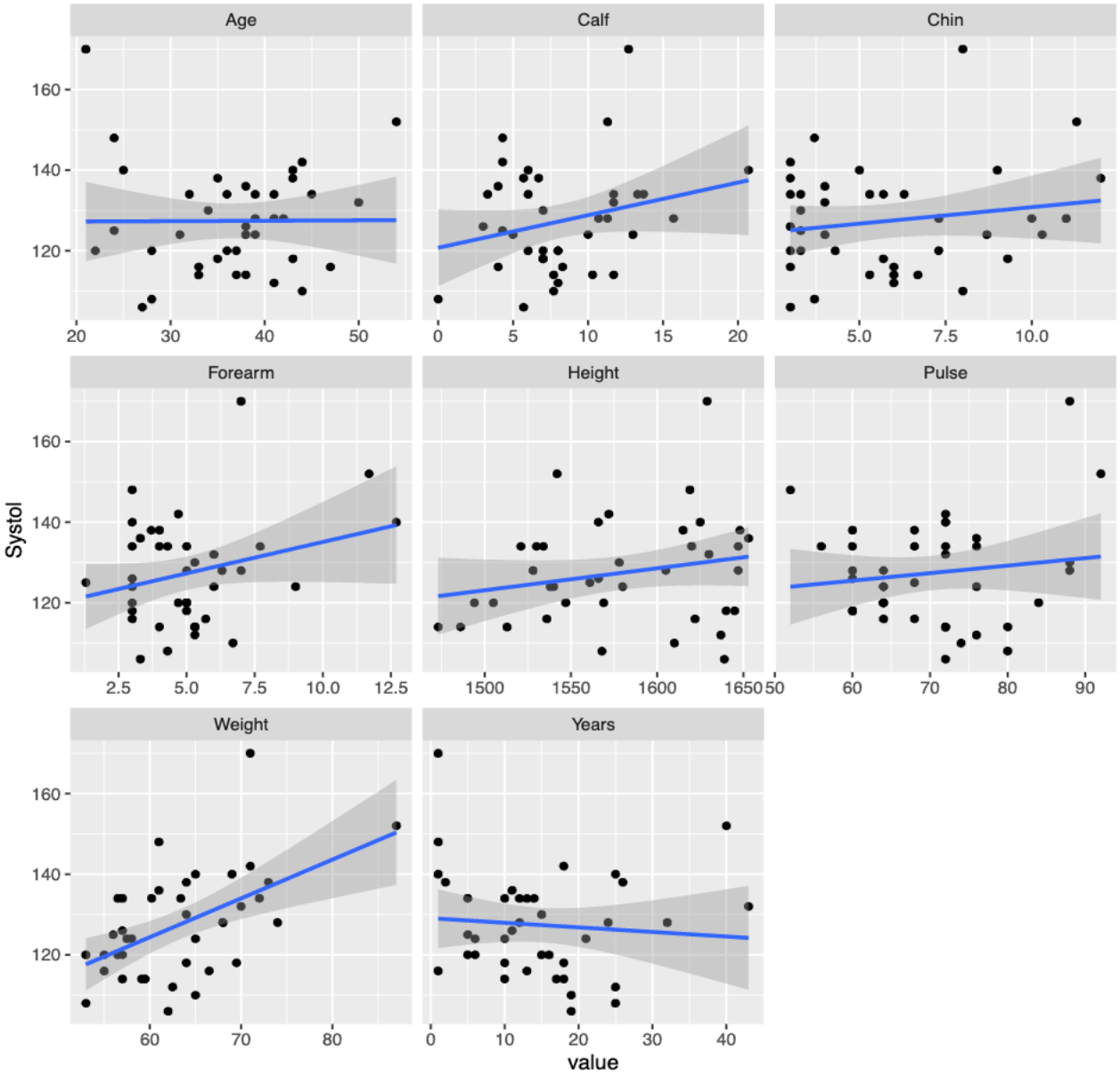
```
blood_pressure %>%
  pivot_longer(cols = Age:Pulse, names_to = "measurement", values_to = "value") %>%
  ggplot() +
    geom_point(mapping = aes(x = value, y = Systol)) +
    facet_wrap(~ measurement, scales = "free_x")
```



ii.

```
blood_pressure %>%
  pivot_longer(cols = Age:Pulse, names_to = "measurement", values_to = "value") %>%
  ggplot() +
    geom_point(mapping = aes(x = value, y = Systol)) +
    facet_wrap(~ measurement, scales = "free_x")+   geom_smooth(mapping = aes(x = value, y = Sy
```

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



Exercise 2

i. Years had a moderate, negative correlation with systol.

ii. I would say the Calf variable had a moderate to strong positive correlation with Systol.

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() %>% select(r.squared, adj.r.squared)
```

r.squared	adj.r.squared
0.0762564	0.0512904

Exercise 6

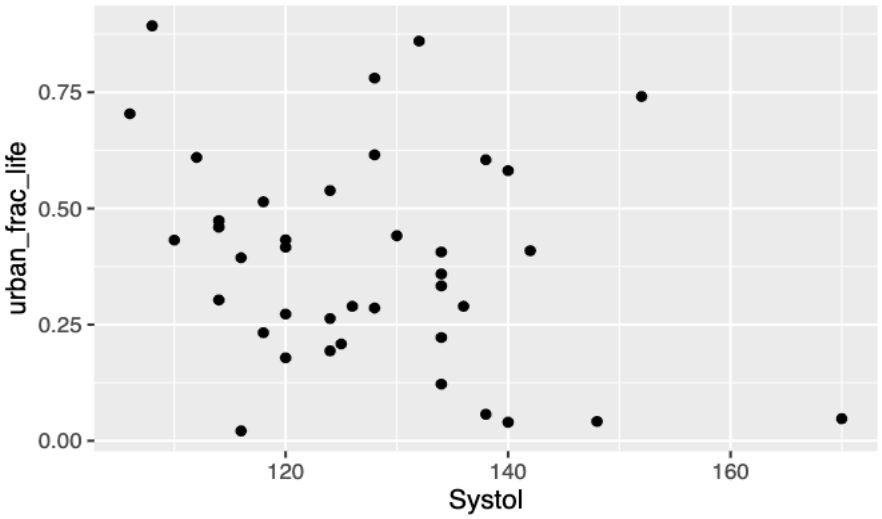
```
systol_urban_frac_df <- blood_pressure_updated %>%
  add_predictions(systol_urban_frac_model) %>%
  add_residuals(systol_urban_frac_model)
```

i. The name of the column that holds response (y) values predicted by the model is "pred".

ii. The name of the column that holds the residuals for each observation is "resid".

Exercise 7

```
ggplot(systol_urban_frac_df) +
  geom_point(mapping = aes(x = Systol, y = urban_frac_life)) +
  geom_abline(slope = -15.8, intercept = 133.5)
```



- Yes, the model does meet the first condition of linearity.

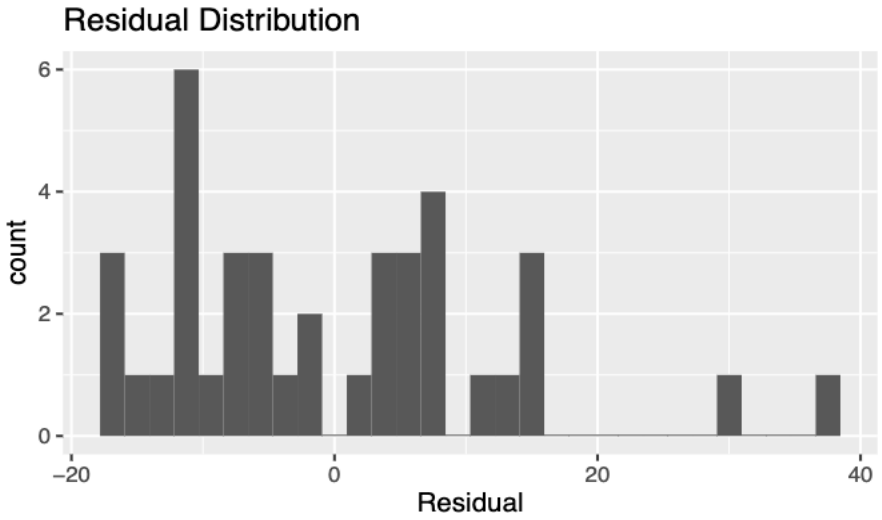
Exercise 8

- It does not look like the variability is reasonably constant all the way along the line. There are many outliers along the way. This means that the predicted value of the error's variability is not constant.

Exercise 9

```
systol_urban_frac_df %>%
  ggplot() +
  geom_histogram(mapping = aes(x = resid)) +
  labs(title = "Residual Distribution", x = "Residual")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



i. The shape in the left is skewed to the right while the middle is skewed to the left. The outer

right is symmetric.

ii. Since this histogram is neither unimodal nor symmetric, it does not meet the nearly normal residuals condition.

Exercise 10

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

```
systol_urban_frac_model %>%
  glance() %>% select(r.squared, adj.r.squared)
```

r.squared	adj.r.squared
0.0762564	0.0512904

```
systol_weight_model %>%
  glance() %>% select(r.squared, adj.r.squared)
```

r.squared	adj.r.squared
0.2718207	0.2521402

Exercise 11

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

Exercise 12