

Statistical analysis of the association between creatinine and weight

By Kieran Arasu

The task

- Primary **predictor** variable is **weight**.
- Primary **outcome** variable is **creatinine levels** (micromoles/L) on admission
- Task is to investigate associations between each variable (except ICU_stay and hosp_stay) and **creatinine levels**.
- **Weight** cannot be excluded from any multivariable regression model.

Variables
Age
Sex
Respiratory
Cardiovascular
Weight
SBP_high
SBP_low
HR
Nurse
Creat
ICU_Stay
Hosp_Stay

Initial analysis

- After some minor data formatting.
- Multivariable analysis of ALL predictor variables (excluding ID, ICU_Stay and hosp_stay.)

	Variables	VIF
1	weight	1.401264
2	age	1.033925
3	sex2Female	1.516968
4	respiratory2Yes	1.127650
5	cardiovascular2Yes	1.124349
6	SBP_high	3.939949
7	SBP_low	3.937077
8	HR	1.017902
9	nurse2Yes	1.062789

```
#data formatting,
hosp2022$sex2 <- factor(hosp2022$sex, labels = c("Male", "Female"))
hosp2022$respiratory2 <- factor(hosp2022$respiratory, labels = c("No", "Yes"))
hosp2022$cardiovascular2 <- factor(hosp2022$cardiovascular, labels = c("No", "Yes"))
hosp2022$nurse2 <- factor(hosp2022$nurse, labels = c("No", "Yes"))
```

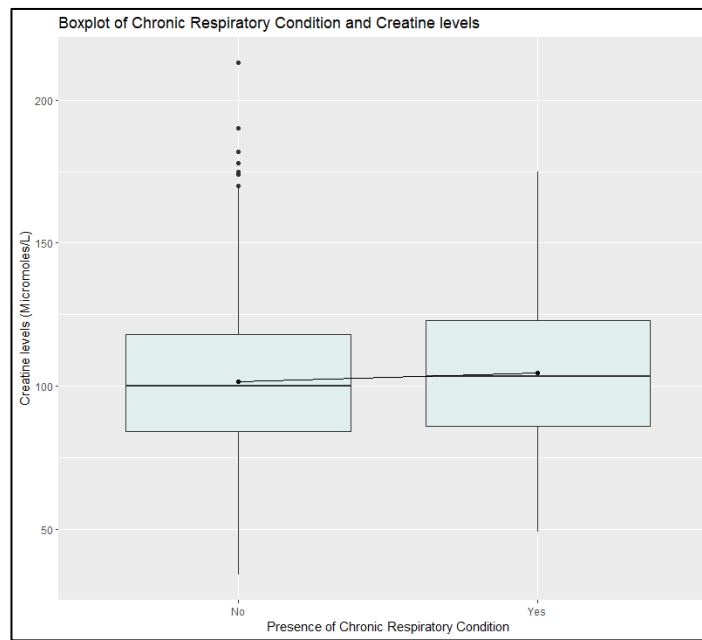
```
Call:
lm(formula = creat ~ weight + age + sex2 + respiratory2 + cardiovascular2 +
    SBP_high + SBP_low + HR + nurse2, data = hosp2022)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.912807   6.493141   2.451  0.0143 *
weight        0.455422   0.034884  13.055 <2e-16 ***
age           1.041230   0.092315  11.279 <2e-16 ***
sex2Female   -28.860929   0.887132 -32.533 <2e-16 ***
respiratory2Yes  0.636269   1.092304   0.583  0.5603
cardiovascular2Yes 0.294462   0.788696   0.373  0.7089
SBP_high      0.022939   0.027392   0.837  0.4025
SBP_low     -0.011425   0.023885  -0.478  0.6325
HR            0.008255   0.014283   0.578  0.5633
nurse2Yes     0.842403   0.836659   1.007  0.3141
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.1 on 1990 degrees of freedom
Multiple R-squared:  0.5531,    Adjusted R-squared:  0.551
F-statistic: 273.6 on 9 and 1990 DF,  p-value: < 2.2e-16
```

Univariate analyses of all Predictor variable (Respiratory and cardiovascular)

- First, predictor variables with no discernable association.
- Multivariable analysis of ALL predictor variables.

Variables
Age
Sex
Respiratory
Cardiovascular
Weight
SBP_high
SBP_low
HR
Nurse
Creat

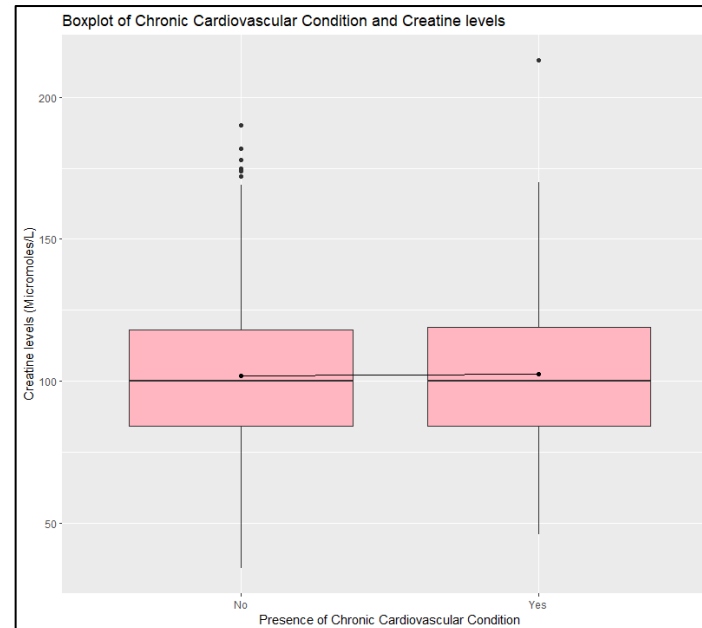


```
Call:
lm(formula = creat ~ respiratory2, data = hosp2022)

Residuals:
    Min       1Q   Median       3Q      Max
-67.651 -18.476  -1.651  16.524 111.349

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   101.6511    0.5802  175.206  <2e-16 ***
respiratory2Yes    2.8244    1.5342   1.841   0.0658 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.02 on 1998 degrees of freedom
Multiple R-squared:  0.001693, Adjusted R-squared:  0.001194
F-statistic: 3.389 on 1 and 1998 DF, p-value: 0.06578
```



```
Call:
lm(formula = creat ~ cardiovascular2, data = hosp2022)

Residuals:
    Min       1Q   Median       3Q      Max
-67.868 -18.367  -1.868  16.633 110.633

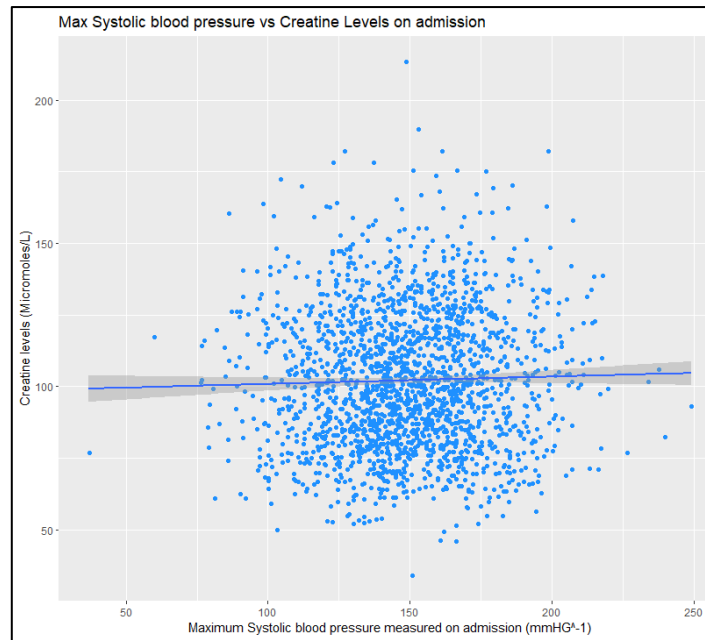
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   101.8680    0.6799  149.823  <2e-16 ***
cardiovascular2Yes    0.4987    1.1103   0.449   0.653
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.04 on 1998 degrees of freedom
Multiple R-squared:  0.0001009, Adjusted R-squared: -0.0003995
F-statistic: 0.2017 on 1 and 1998 DF, p-value: 0.6534
```

Univariate analyses of all Predictor variable (SBP_High and SBP_Low)

- First, predictor variables with no discernable association.
- No significant relationship found between systolic blood pressure (min and max) and creatinine levels.

Variables
Age
Sex
Respiratory
Cardiovascular
Weight
SBP_high
SBP_low
HR
Nurse
Creat

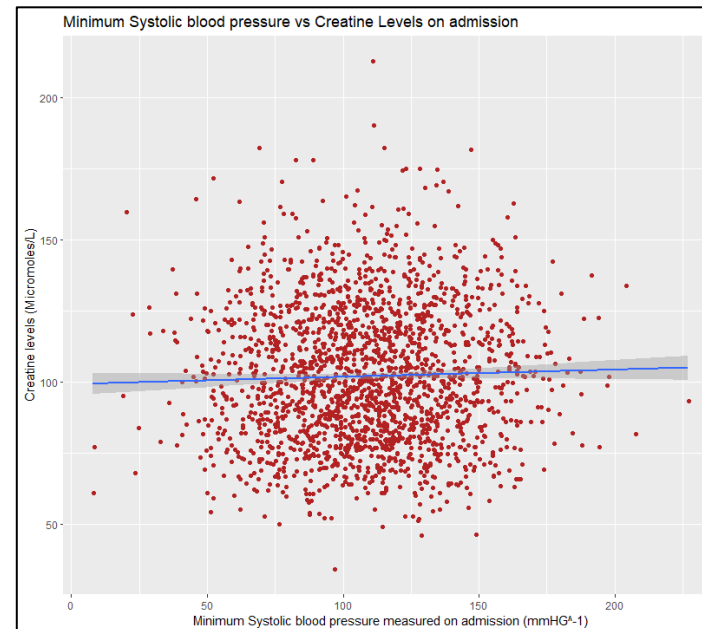


```
Call:
lm(formula = creat ~ SBP_high, data = hosp2022)

Residuals:
    Min       1Q   Median       3Q      Max
-68.112 -18.317  -1.893  16.568 110.940

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  98.23210    3.11094   31.576  <2e-16 ***
SBP_high      0.02569    0.02059    1.248   0.212
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.03 on 1998 degrees of freedom
Multiple R-squared:  0.0007784, Adjusted R-squared:  0.0002783
F-statistic: 1.557 on 1 and 1998 DF, p-value: 0.2123
```



```
Call:
lm(formula = creat ~ SBP_low, data = hosp2022)

Residuals:
    Min       1Q   Median       3Q      Max
-67.754 -18.270  -1.995  16.648 110.890

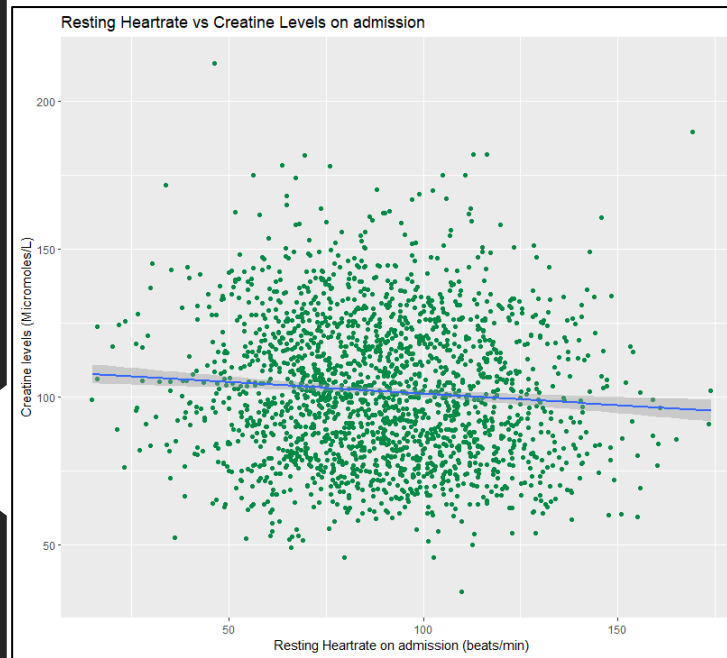
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  99.28618    2.02753   48.969  <2e-16 ***
SBP_low      0.02544    0.01796    1.416   0.157
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.03 on 1998 degrees of freedom
Multiple R-squared:  0.001003, Adjusted R-squared:  0.0005029
F-statistic: 2.006 on 1 and 1998 DF, p-value: 0.1569
```

Univariate analyses of all Predictor variable (HR and Nurse Program)

- HR and Nurse have significant relationship with creatinine.
- If they are independent predictors of creatinine, they should be included in the final model.

Variables
Age
Sex
Respiratory
Cardiovascular
Weight
SBP_high
SBP_low
HR
Nurse
Creat

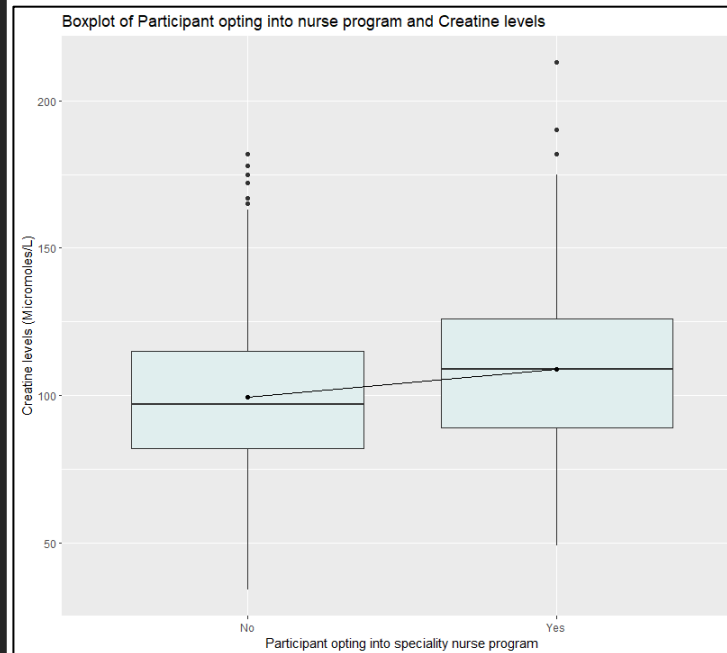


```
Call:
lm(formula = creat ~ HR, data = hosp2022)

Residuals:
    Min       1Q   Median       3Q      Max
-66.446 -17.876  -1.966   16.723  107.563

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  109.02398    1.95685   55.714  < 2e-16 ***
HR          -0.07799    0.02106   -3.703  0.000219 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.96 on 1998 degrees of freedom
Multiple R-squared:  0.006815, Adjusted R-squared:  0.006318 
F-statistic: 13.71 on 1 and 1998 DF, p-value: 0.000219
```



```
Call:
lm(formula = creat ~ nurse2, data = hosp2022)

Residuals:
    Min       1Q   Median       3Q      Max
-65.537 -17.537  -1.537   16.463  104.119

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   99.5366    0.6195  160.671  < 2e-16 ***
nurse2Yes      9.3446    1.1933   7.831  7.82e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.68 on 1998 degrees of freedom
Multiple R-squared:  0.02978, Adjusted R-squared:  0.02929 
F-statistic: 61.32 on 1 and 1998 DF, p-value: 7.815e-15
```


Univariate analyses of all Predictor variable (Age and Sex)

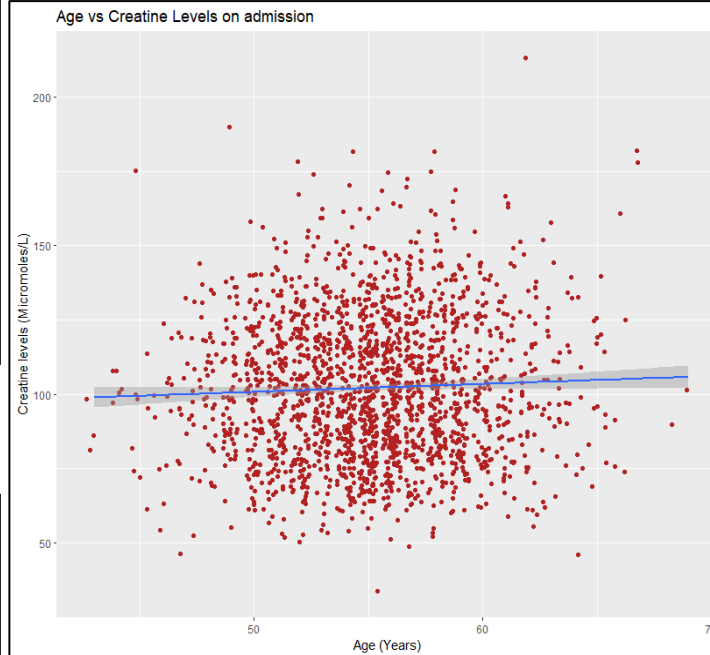
- No significant relationship for Age
- Significant relationship for Sex
- However, very significant relationship was noted earlier.

```
Call:
lm(formula = creat ~ weight + age + sex2 + respiratory2 + cardiovascular2 +
    SBP_high + SBP_low + HR + nurse2, data = hosp2022)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-46.935 -10.626  -1.198   9.427  81.045
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.912807   6.493141    2.451  0.0143 *
weight       0.455422   0.034884   13.055 <2e-16 ***
age          1.041230   0.092315   11.279 <2e-16 ***
sex2Female  -28.860929   0.887132  -32.533 <2e-16 ***
respiratory2Yes  0.636269   1.092304    0.583  0.5603
cardiovascular2Yes 0.294462   0.788696    0.373  0.7089
SBP_high     0.022939   0.027392    0.837  0.4025
SBP_low     -0.011425   0.023885   -0.478  0.6325
HR           0.008255   0.014283    0.578  0.5633
nurse2Yes    0.842403   0.836659    1.007  0.3141
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 16.1 on 1990 degrees of freedom
Multiple R-squared:  0.5531,    Adjusted R-squared:  0.551
F-statistic: 273.6 on 9 and 1990 DF,  p-value: < 2.2e-16
```

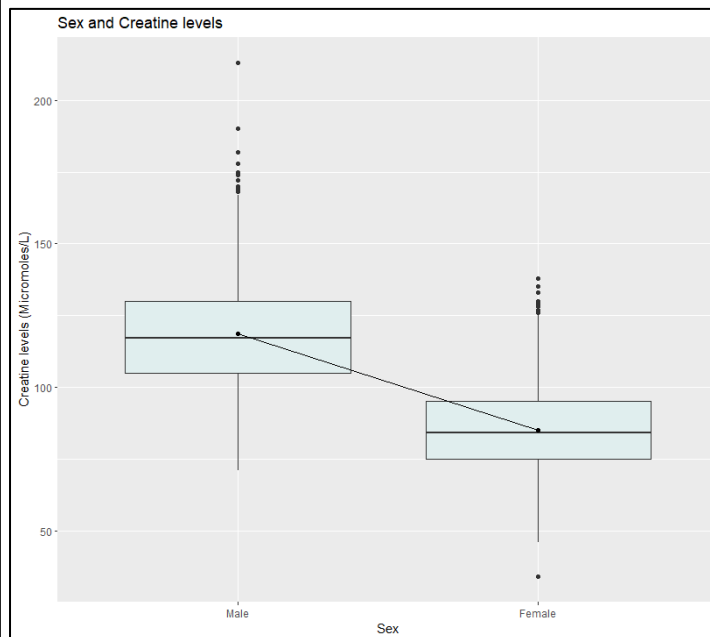


```
Call:
lm(formula = creat ~ age, data = hosp2022)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-68.026 -18.291  -1.497   16.768  109.120
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  87.4596    7.4810   11.691 <2e-16 ***
age          0.2648    0.1354    1.956  0.0506 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 24.02 on 1998 degrees of freedom
Multiple R-squared:  0.001911,    Adjusted R-squared:  0.001412
F-statistic: 3.826 on 1 and 1998 DF,  p-value: 0.0506
```



```
lm(formula = creat ~ sex2, data = hosp2022)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-51.049 -11.526  -1.049   10.474   94.474
```

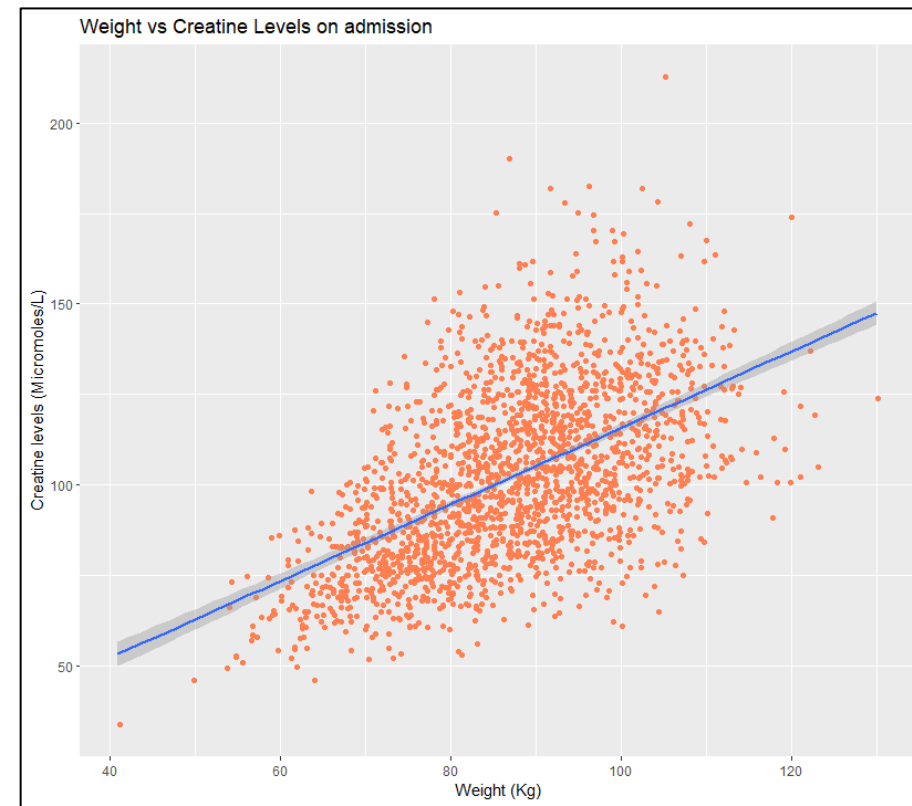
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  118.5256    0.5412   219.02 <2e-16 ***
sex2Female  -33.4768    0.7715  -43.39 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 17.25 on 1998 degrees of freedom
Multiple R-squared:  0.4852,    Adjusted R-squared:  0.4849
F-statistic: 1883 on 1 and 1998 DF,  p-value: < 2.2e-16
```

Univariate analyses of all Predictor variable (Weight)

- Very significant relationship
- No homoscedasticity

Variables
Age
Sex
Respiratory
Cardiovascular
Weight
SBP_high
SBP_low
HR
Nurse
Creat



```
Call:
lm(formula = creat ~ weight, data = hosp2022)

Residuals:
    Min       1Q   Median       3Q      Max
-55.037 -14.185  -2.562  12.828  91.904

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.97652    3.25776   3.062  0.00223 **
weight       1.05828    0.03708  28.542 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

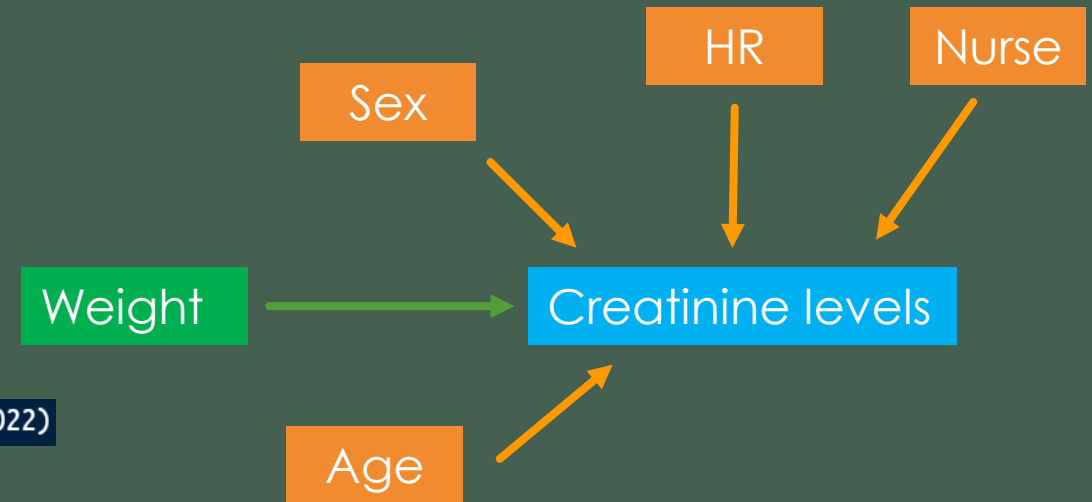
Residual standard error: 20.26 on 1998 degrees of freedom
Multiple R-squared:  0.2896,    Adjusted R-squared:  0.2893 
F-statistic: 814.6 on 1 and 1998 DF,  p-value: < 2.2e-16
```


Post-Univariate analyses

- Primary **predictor** variable is **weight**.
- Primary **outcome** variable is **creatinine levels** (micromoles/L) on admission
- After investigating each predictor variable, the following variables will be included into an initial regression model.

```
lm(formula = creat ~ weightc + agec + sex2 + HRc + nurse2, data = hosp2022)
```

	Variables	Tolerance	VIF
1	weightc	0.7150674	1.398470
2	agec	0.9675732	1.033513
3	sex2Female	0.6614188	1.511901
4	HRc	0.9828138	1.017487
5	nurse2Yes	0.9423922	1.061129



Initial Model

- The model provides the following results

```
lm(formula = creat ~ weightc + agec + sex2 + HRc + nurse2, data = hosp2022)

Residuals:
    Min       1Q   Median       3Q      Max
-47.559 -10.703  -1.157   9.411  81.077

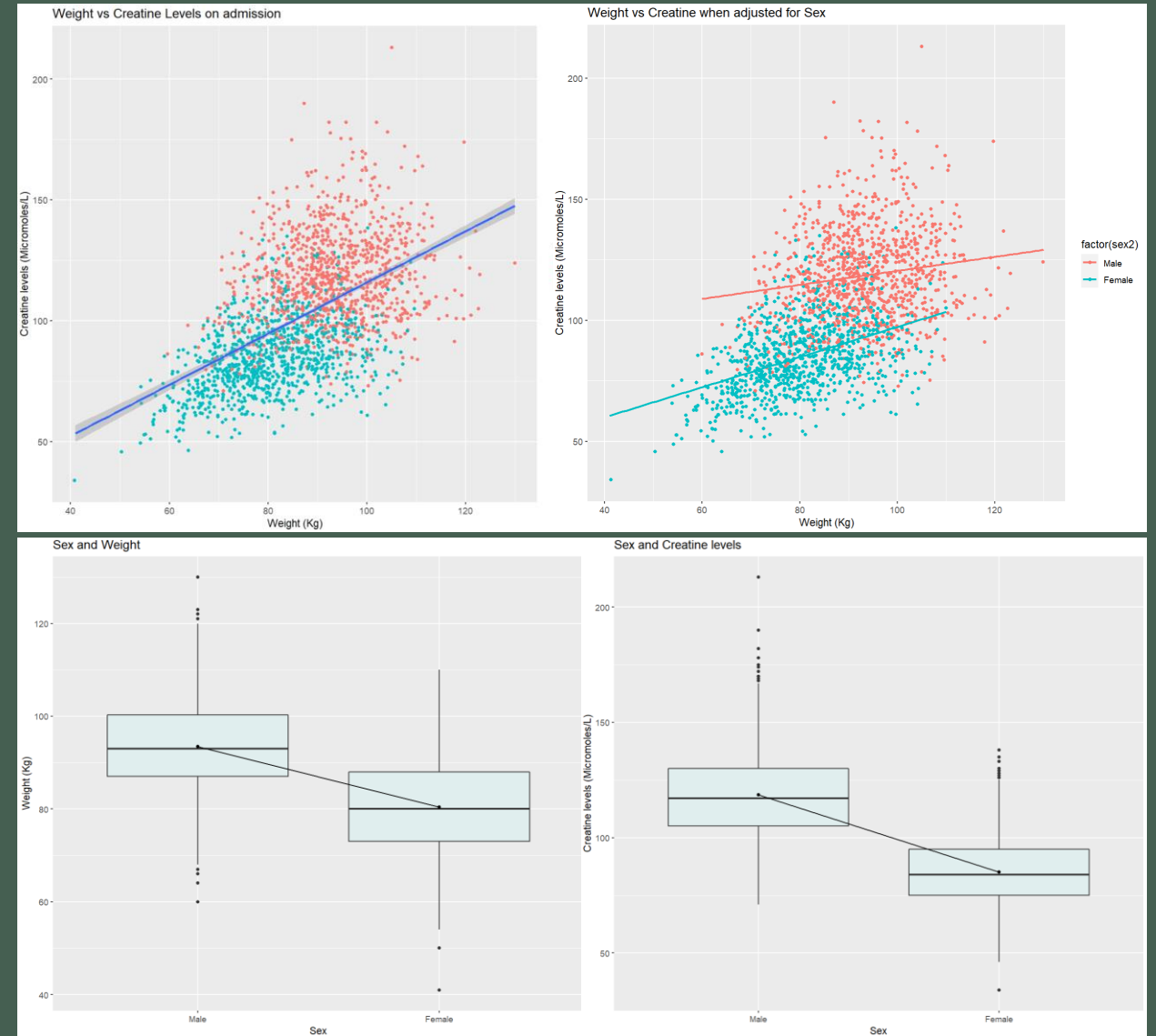
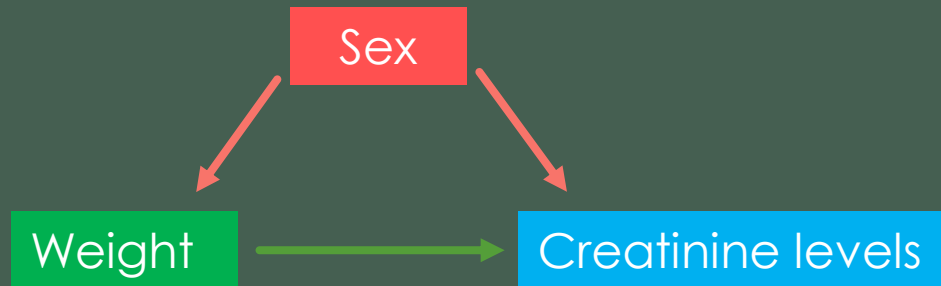
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  116.020264   0.640040  181.270  <2e-16 ***
weightc       0.456902   0.034829   13.119  <2e-16 ***
agec          1.042740   0.092241   11.305  <2e-16 ***
sex2Female   -28.852277   0.885115  -32.597  <2e-16 ***
HRc           0.008238   0.014271    0.577    0.564
nurse2Yes     0.853641   0.835501    1.022    0.307
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.09 on 1994 degrees of freedom
Multiple R-squared:  0.5527,    Adjusted R-squared:  0.5516
F-statistic: 492.8 on 5 and 1994 DF,  p-value: < 2.2e-16
```

- As can be noted, weight, sex and age still have a significant relationship with creatinine levels.
- HR and Nurse variable have no significant relationship.

Building on the initial Model

- Sex is actively confounding weight (and other variables) (associated with both weight and creatinine levels)
- In the dataset, Men are both heavier and take more creatinine.
- Therefore it should be kept in the regression model and be adjusted for.



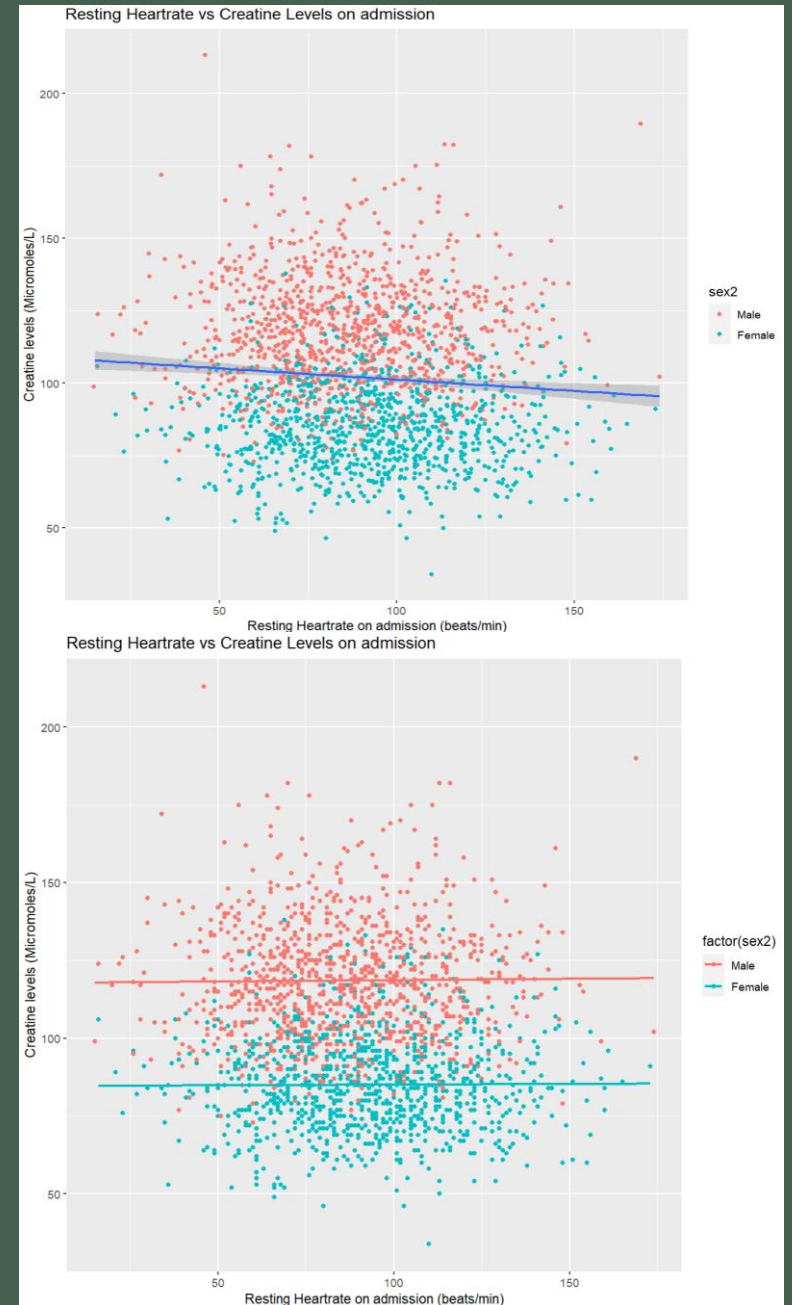
Variable Pruning (HR)

- When adjusting for sex, the HR predictor variable has no significant relationship with creatinine intake.

```
lm(formula = creat ~ HRC + sex2, data = hosp2022)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 118.321527   0.719959  164.34  <2e-16 ***
HRC          0.006574   0.015293    0.43   0.667
sex2Female   -33.519740   0.778109  -43.08  <2e-16 ***
```

- It also has no significant interaction with weight in the initial model.

```
lm(formula = creat ~ weightc + agec + sex2 + HRC + nurse2 + weightc *
    HRC, data = hosp2022)
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.157e+02  7.720e-01 149.926  <2e-16 ***
weightc     4.769e-01  5.335e-02  8.939   <2e-16 ***
agec        1.045e+00  9.237e-02 11.313  <2e-16 ***
sex2Female  -2.885e+01  8.853e-01 -32.591  <2e-16 ***
HRC         7.826e-03  1.430e-02  0.547   0.584
nurse2Yes   8.607e-01  8.358e-01  1.030   0.303
weightc:HRC -5.862e-04  1.185e-03 -0.495   0.621
```



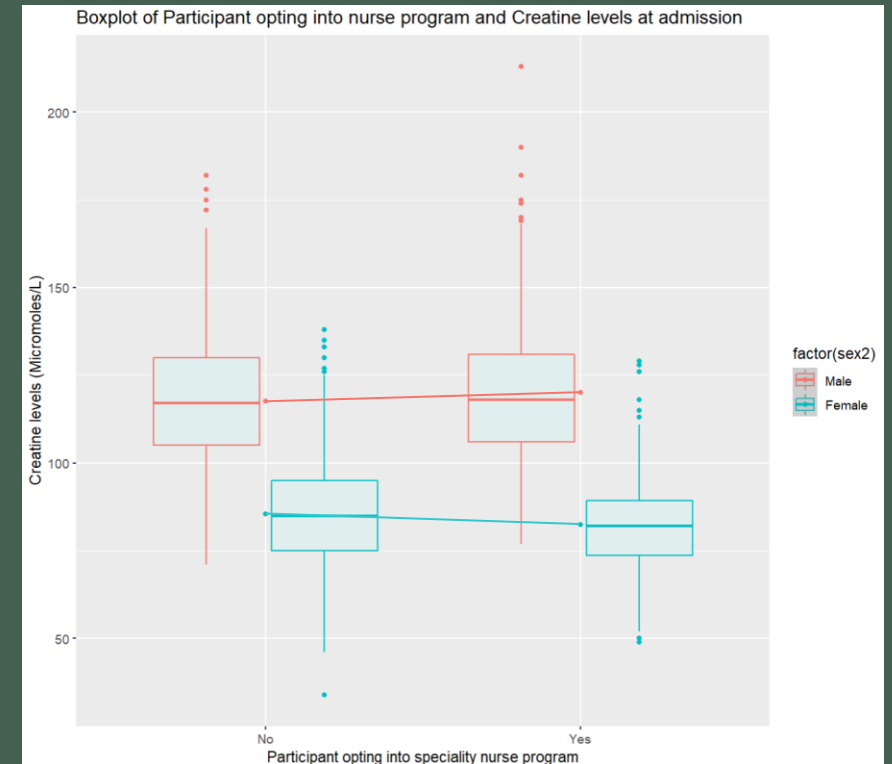
Variable Pruning (Nurse)

- When adjusting for sex, the nurse predictor variable has no significant relationship with creatinine levels.

```
lm(formula = creat ~ nurse2 + sex2, data = hosp2022)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  118.3670     0.6359  186.126  <2e-16 ***
nurse2Yes     0.4252     0.8950   0.475    0.635
sex2Female   -33.3873     0.7943 -42.033  <2e-16 ***
```

- It also has no significant interaction with weight in the initial model.

```
lm(formula = creat ~ weightc + agec + sex2 + HRc + nurse2 + weightc *
    nurse2, data = hosp2022)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  115.750836    0.771662  150.002  <2e-16 ***
weightc      0.428451    0.039959   10.722  <2e-16 ***
agec         1.038224    0.092268   11.252  <2e-16 ***
sex2Female   -28.900681    0.885498  -32.638  <2e-16 ***
HRc          0.007974    0.014269   0.559    0.576
nurse2Yes    0.706493    0.841403   0.840    0.401
weightc:nurse2Yes 0.095331    0.065696   1.451    0.147
```



Variable adjustments (age)

- Apparent association between age and weight.

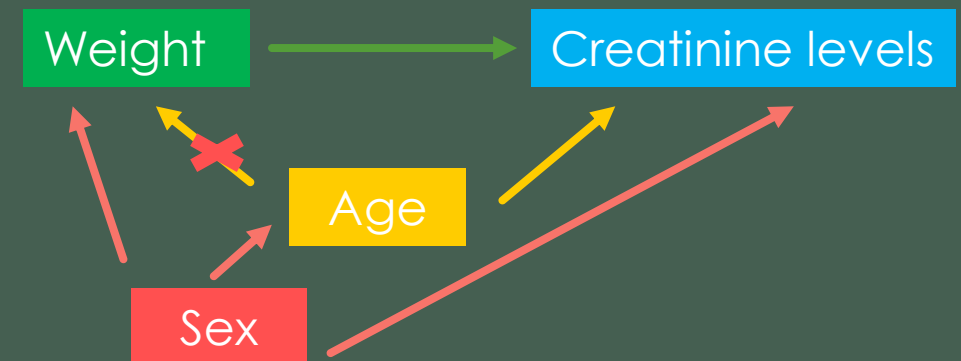
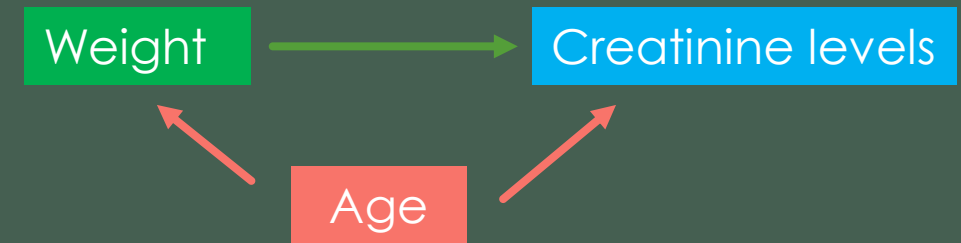
```
lm(formula = weight ~ age, data = hosp2022)
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 102.58895    3.79193   27.05 < 2e-16 ***
age         -0.28273    0.06863   -4.12 3.95e-05 ***
```

- This would mean age is a confounder for weight, however after adjusting for sex.

```
lm(formula = weight ~ age + sex2, data = hosp2022)
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  93.001197    3.239695  28.707 <2e-16 ***
age           0.007791    0.059233   0.132  0.895
sex2Female   -13.053997    0.469933 -27.778 <2e-16 ***
```

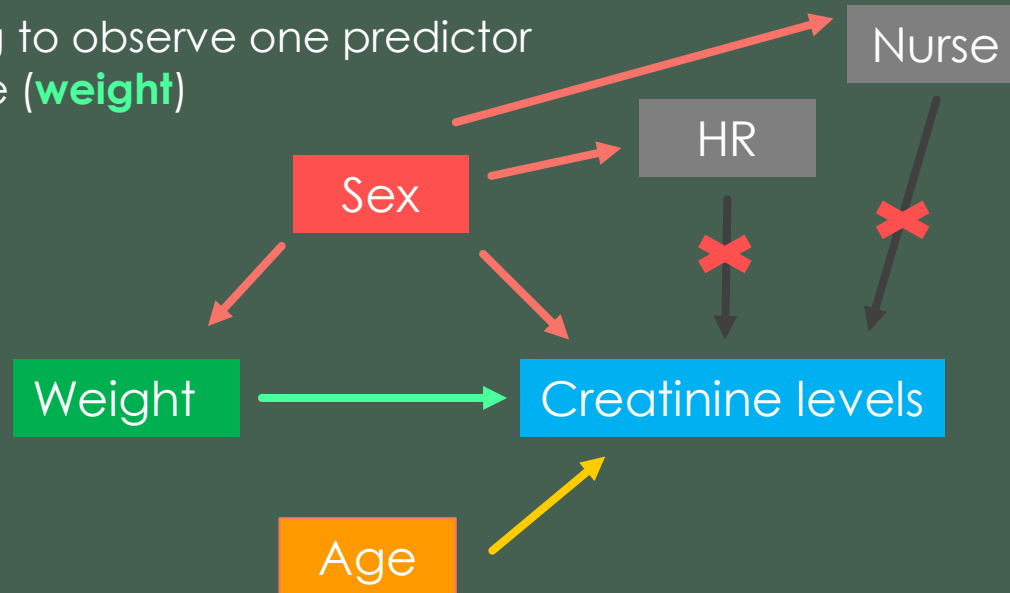
- There is no association. This means age is an independent predictor of creatinine levels and should be included in the model.

```
lm(formula = creat ~ sex2 + age, data = hosp2022)
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  61.79940    5.25231  11.77 <2e-16 ***
sex2Female   -34.93702    0.76187 -45.86 <2e-16 ***
age           1.04239    0.09603  10.86 <2e-16 ***
```



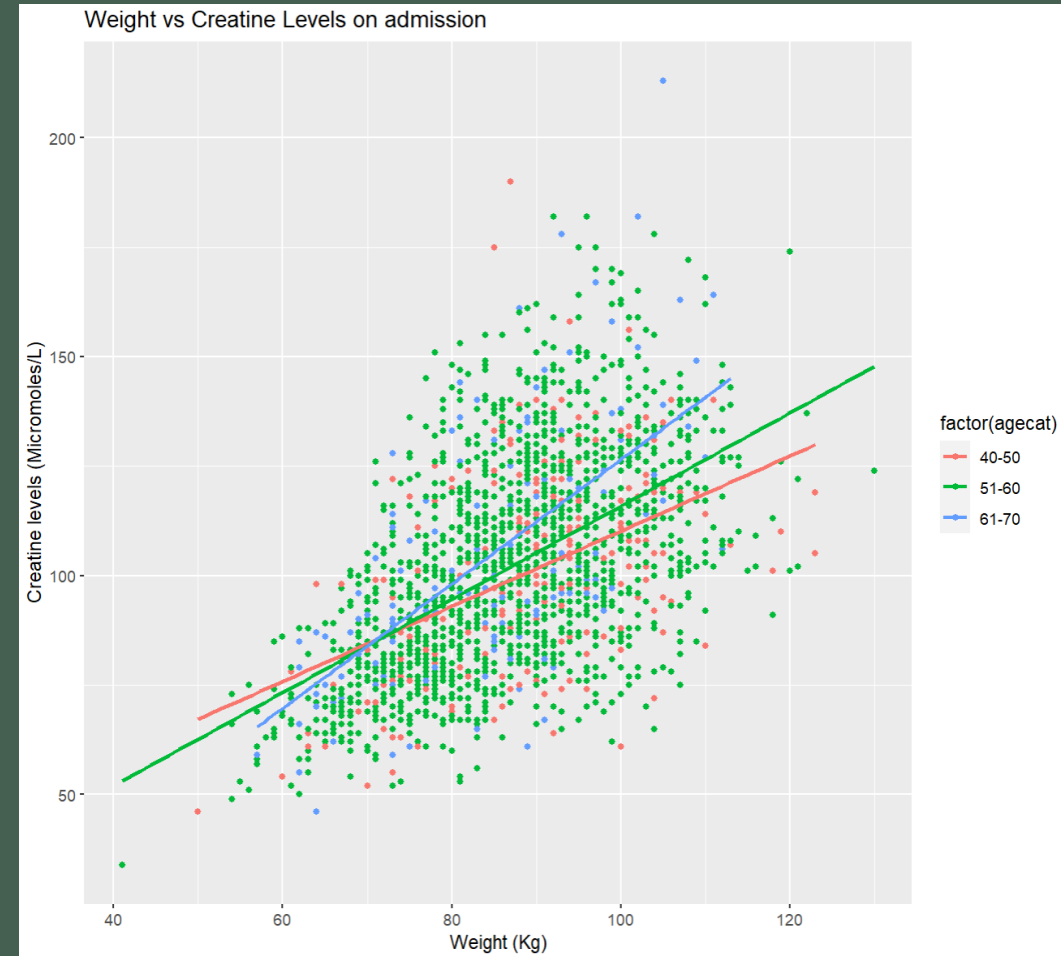
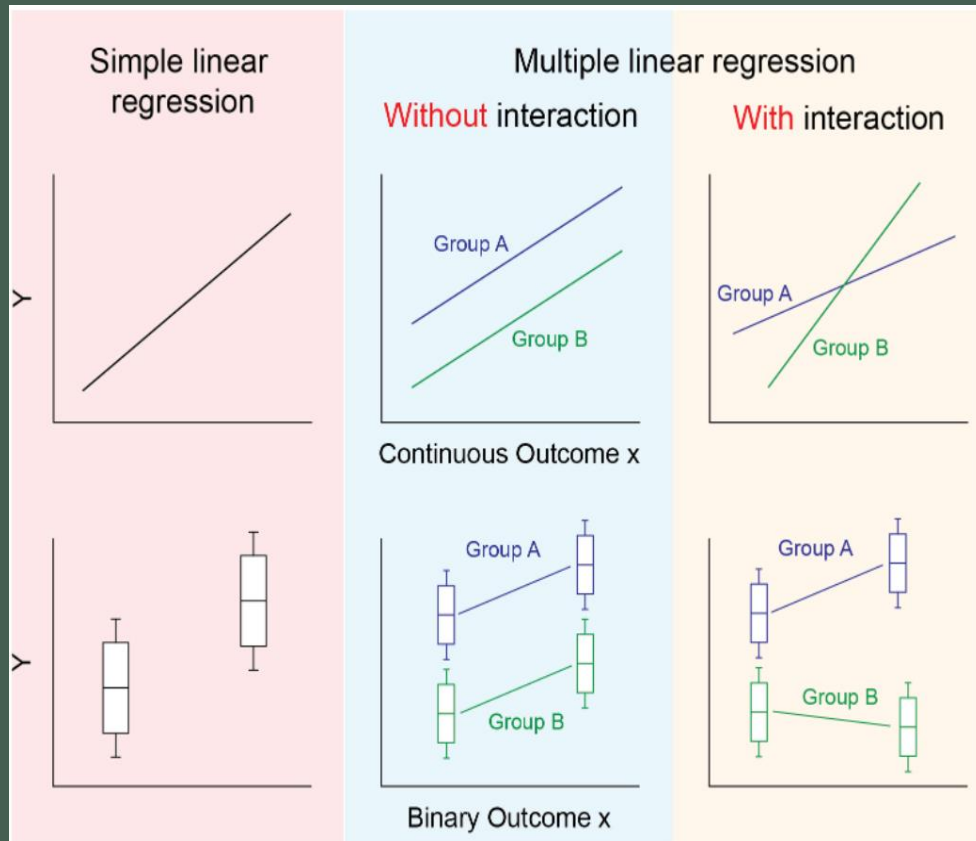
Building on the initial Model

- Parsimonious Approach
- Working to observe one predictor variable (**weight**)

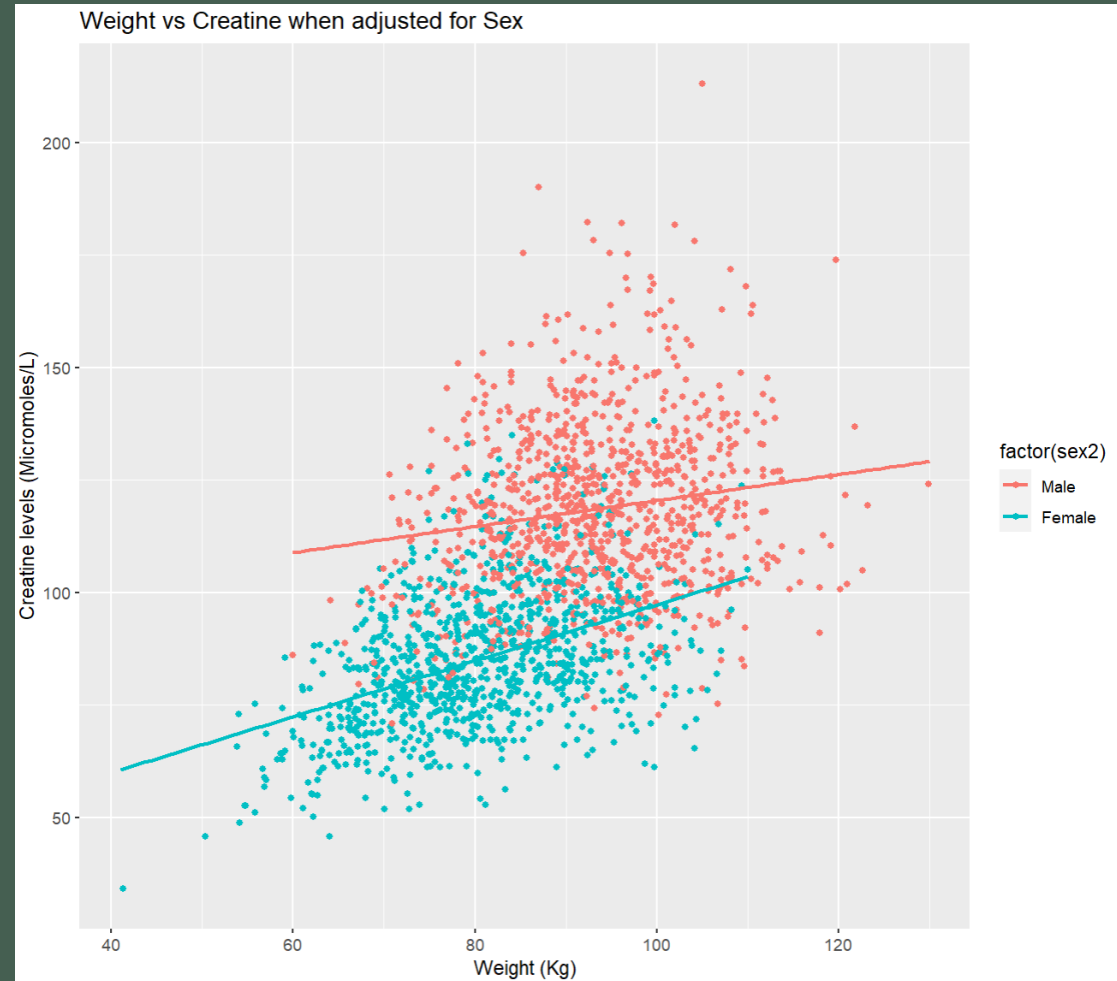
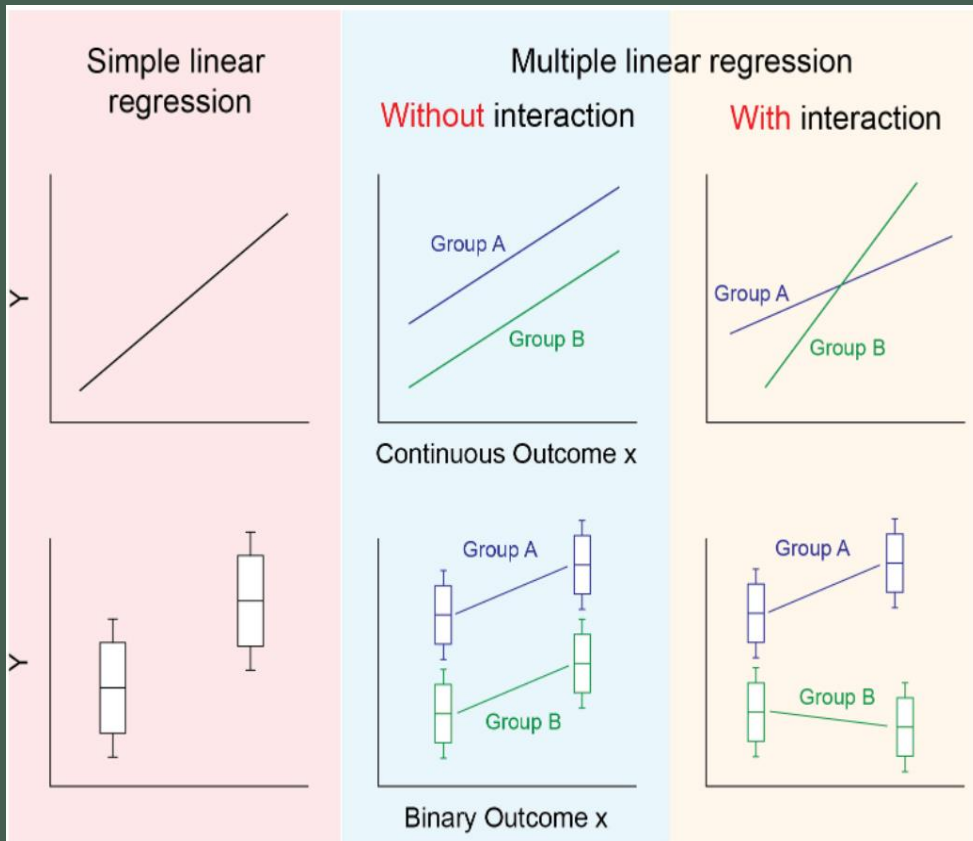


- A spurious association was generated for both HR and Nurse due to sex acting as a confounder.
- No interaction with weight therefore, they can be removed and not be adjusted for.

Interaction term



Interaction term



Final Model with interaction terms

```
lm(formula = creat ~ weightc + sex2 + agec + weightc * sex2 +  
    weightc * agec, data = hosp2022)
```

Residuals:

Min	1Q	Median	3Q	Max
-46.825	-10.626	-1.162	9.105	80.112

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	117.406747	0.595680	197.097	< 2e-16	***
weightc	0.300870	0.049462	6.083	1.41e-09	***
sex2Female	-28.952324	0.854299	-33.890	< 2e-16	***
agec	1.043665	0.091452	11.412	< 2e-16	***
weightc:sex2Female	0.307998	0.069817	4.411	1.08e-05	***
weightc:agec	0.023191	0.007522	3.083	0.00208	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.97 on 1994 degrees of freedom
Multiple R-squared: 0.5598, Adjusted R-squared: 0.5587
F-statistic: 507.1 on 5 and 1994 DF, p-value: < 2.2e-16

	2.5 %	97.5 %
(Intercept)	116.238527459	118.57496705
weightc	0.203868278	0.39787230
sex2Female	-30.627735869	-27.27691254
agec	0.864313454	1.22301653
weightc:sex2Female	0.171074754	0.44492054
weightc:agec	0.008439499	0.03794278

Variables

Age*

Sex*

Respiratory

Cardiovascular

Weight

SBP_high

SBP_low

HR

Nurse

Creat

ICU_Stay

Hosp_Stay

Model Comparison

Initial Model: $Creat = \beta_0 + \beta_1(\text{weight}) + \beta_2(\text{age}) + \beta_3(\text{Sex}) + \beta_4(\text{HR}) + \beta_5(\text{Nurse})$

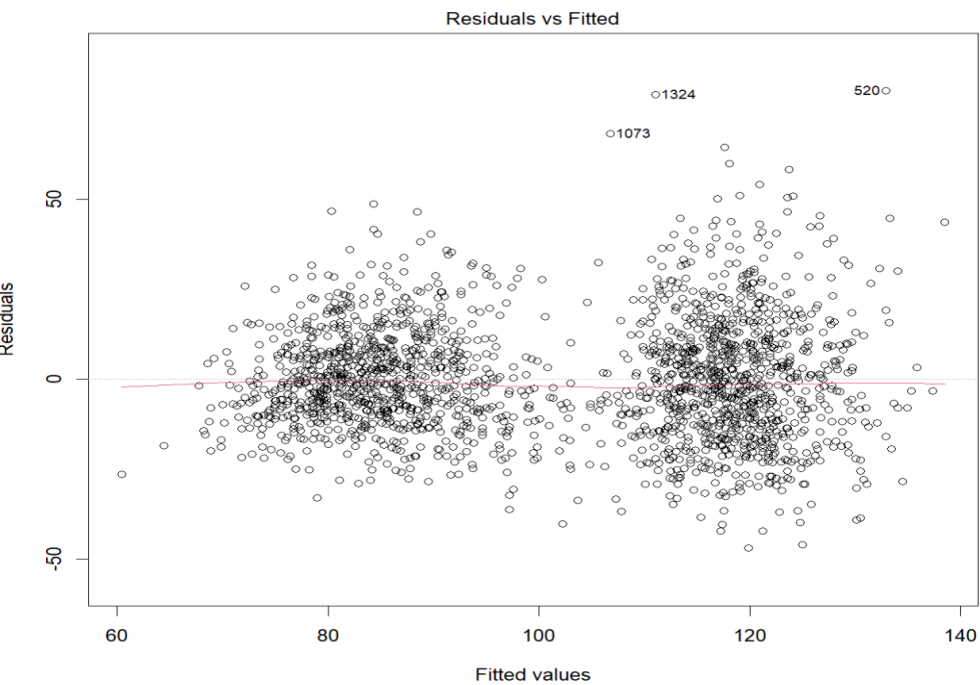
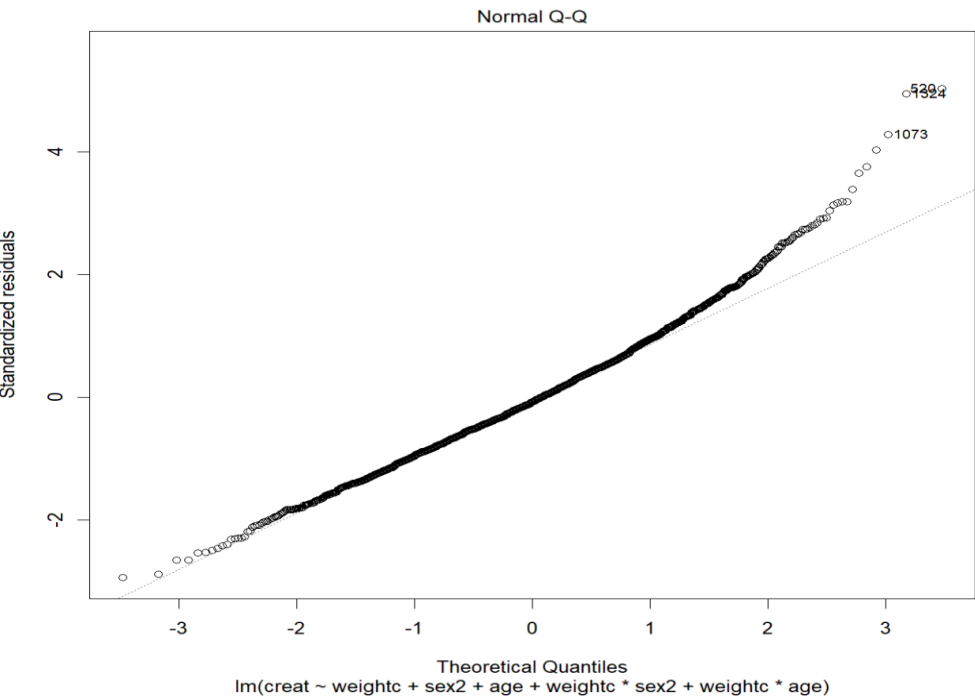
Final Model: $Creat = \beta_0 + \beta_1(\text{weight}) + \beta_2(\text{age}) + \beta_3(\text{Sex}) + \beta_4(\text{weight} * \text{age}) + \beta_5(\text{weight} * \text{sex})$

Final Model (Rearranged): $Creat = \beta_0 + \beta_2(\text{age}) + \beta_3(\text{Sex}) + (\text{weight})(\beta_1 + \beta_4(\text{age}) + \beta_5(\text{sex}))$

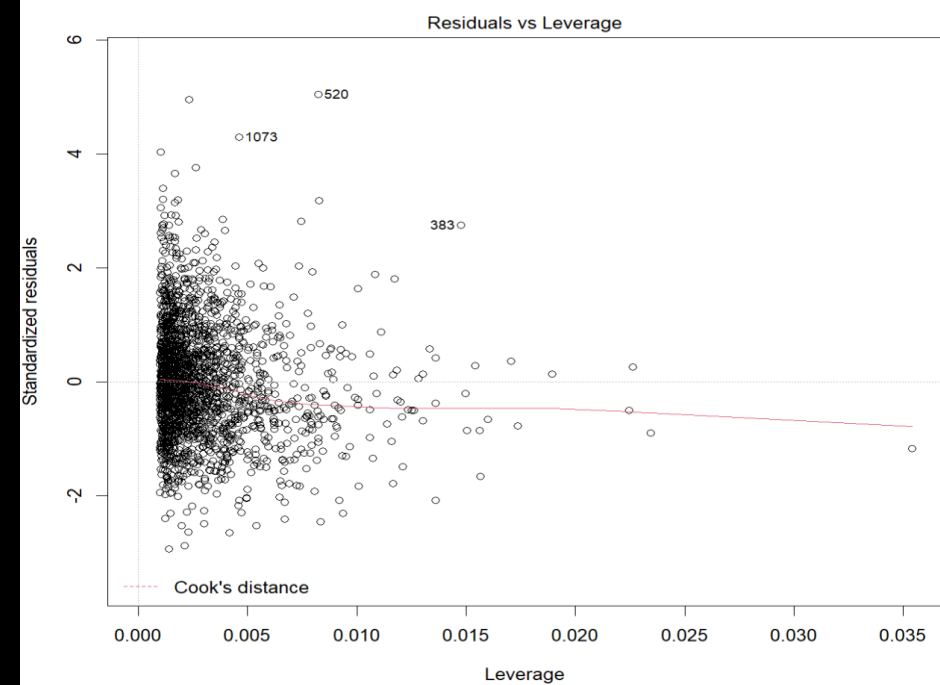
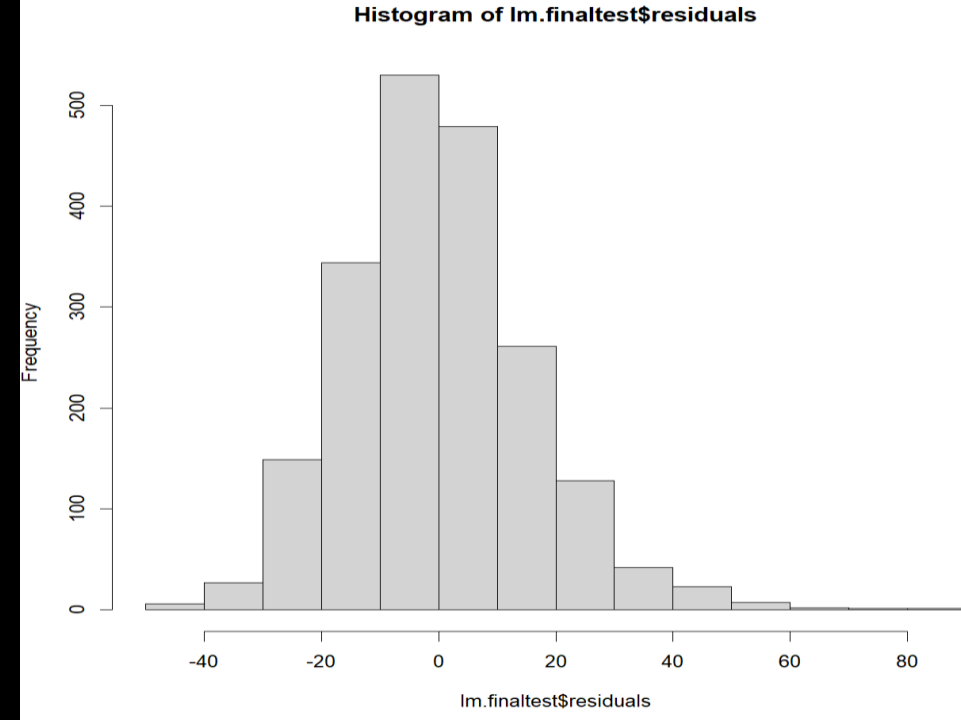
Model	Adj. R^2	AIC	BIC
Initial	0.5516	16797.56	16836.77
Final	0.5587	16765.62	16804.82

Higher values for R^2 means the model is a better fit.

Lower values for both AIC and BIC are indicative of a better model.



- Assumptions test show homoscedasticity and linearity
- Residuals curve upward in QQ plot but large sample size so it isn't as critical.
- Outliers not extreme enough to have significant leverage.
- Overall, we will pick model 3.



Summary of results

Final Model: $Creat = 117.40 - 28.95(Sex) + 1.044(age) + weight(0.30 + 0.308(sex) + 0.023(age))$

	Coefficients (LCI – UCI)	Std.Error	t-Value	p-value
β	117.40 (116.24, 118.57)	0.596	197.097	<0.0001***
Weightc (β_1)	0.30 (0.20, 0.40)	0.05	-2.370	<0.0001***
Sex (β_2)	-28.95 (-30.63, -27.28)	0.85	-33.890	<0.0001***
Age (β_3)	1.044 (0.86, 1.22)	0.09	11.412	<0.0001***
Weight*Sex (β_4)	0.308 (0.17, 0.44)	0.07	4.411	<0.0001***
Weight*Age (β_5)	0.023 (0.01-0.03)	0.008	3.083	<0.002**

R^2 : 0.5598

This model accounts for 55.98%
of variability of creatinine

- In summary, there is a very significant association between weight and creatinine levels (p-value <0.0001).
- For every 1Kg increase in weight. The mean creatinine levels on admission increase by 0.30 micromoles/L.
- Weight also has significant interactions with sex and age to increase this effect.
- For every one-year increase of age, this effect increases by 0.023. Females also increase this effect of weight by 0.308.
- A male with an average weight and age would have 117.40 micromole/L on admission.