

Exam 1 Review

Grinnell College

February 19, 2024

Review

- ▶ Hwk 14 and 20
- ▶ Plots for variables
- ▶ Tables and plots
- ▶ Correlation
- ▶ Regression
 - ▶ Simple linear (quantitative)
 - ▶ Simple linear (categorical)
 - ▶ Multivariate (quant + categorical)

NYC Flight Data

Below is a table recording flight data departing NYC in 2013

	Newark	JFK	LGA	Sum
American Airlines	350	1388	1450	3188
United	4559	440	771	5770
Southwest	631	0	630	1261
Sum	5540	1828	2851	10219

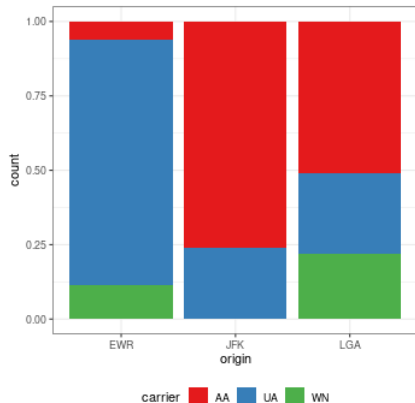
NYC Flight Data

	Newark	JFK	LGA
American	0.06	0.75	0.50
United	0.82	0.25	0.27
Southwest	0.12	0	0.23

- ▶ Conditioned on?
- ▶ Type of plot?
- ▶ What is predictor what is explanatory?

NYC Flight Data

	Newark	JFK	LGA
American	0.06	0.75	0.50
United	0.82	0.25	0.27
Southwest	0.12	0	0.23



Odds Ratio – Police

Are signs of mental illness associated with increased aggression?

	Attack Level	
Mental Illness	Attack	Other
Yes	771	444
No	2,839	1,278

Odds Ratio – Surgery

In the 1860s, Joseph Lister conducted landmark experiment to investigate benefits of sterile technique in surgery:

- ▶ Wash hands
- ▶ Wear clean gloves
- ▶ Disinfect surgical instruments with carbolic acid

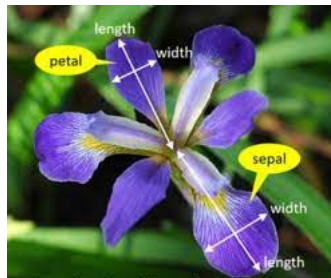
Environment	Survived	
	Yes	No
Sterile	34	6
Control	19	16

Risk factors for breast cancer

In one study, CDC tracked 6,168 women in hopes of finding risk factors for breast cancer. In particular, considered age at which woman gave birth to first child:

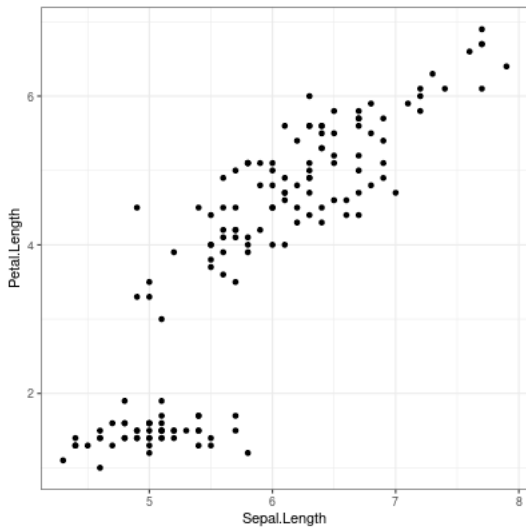
Age at 1st birth	Breast Cancer	
	Yes	No
Before age 25	4475	65
25 or older	1597	31

Iris data



```
1 > head(iris)
2   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
3 1           5.1           3.5           1.4           0.2   setosa
4 2           4.9           3.0           1.4           0.2   setosa
5 3           4.7           3.2           1.3           0.2   setosa
6 4           4.6           3.1           1.5           0.2   setosa
7 5           5.0           3.6           1.4           0.2   setosa
8 6           5.4           3.9           1.7           0.4   setosa
```

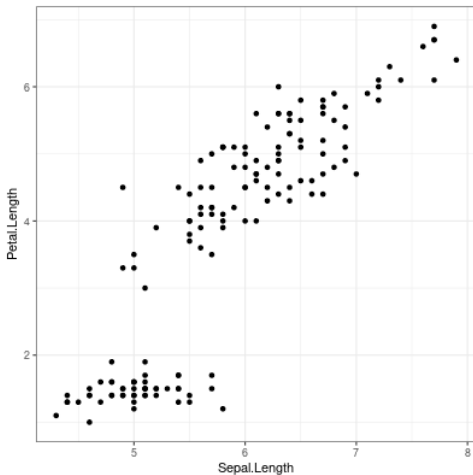
Correlation



Correlation

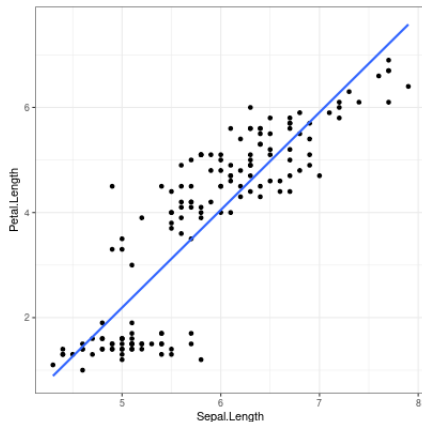
and Regression

$$r = 0.87$$

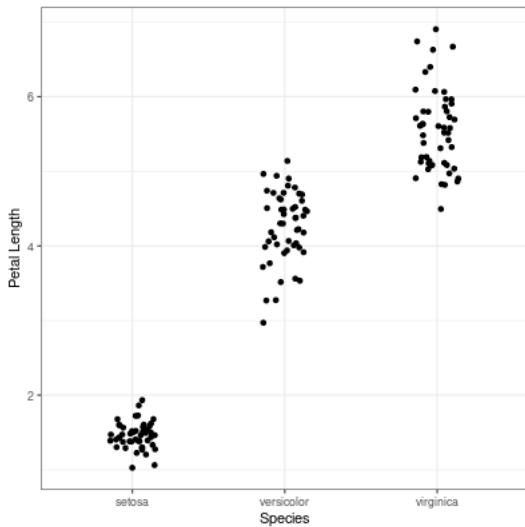


Iris SLR

```
1 > lm(Petal.Length ~ Sepal.Length, iris)
2
3 Coefficients:
4 (Intercept) Sepal.Length
5      -7.10      1.86
```



Categorical



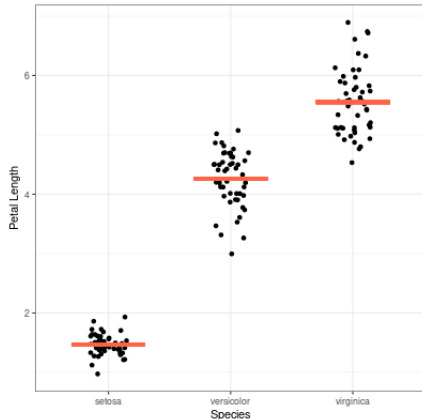
Categorical

```
1 > lm(Petal.Length ~ Species, iris)
```

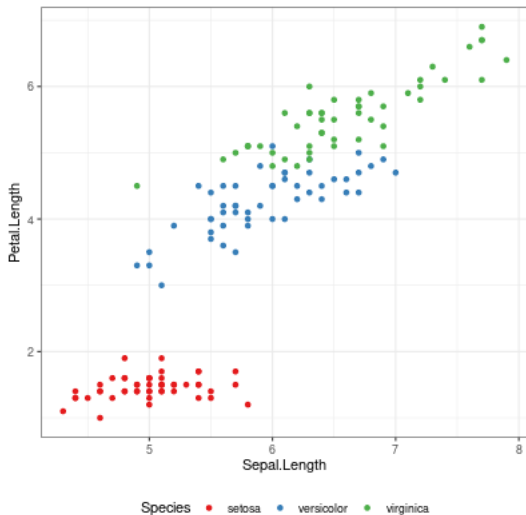
```
2
```

```
3 Coefficients:
```

```
4      (Intercept)  Speciesversicolor  Speciesvirginica
5              1.46                2.80                4.09
```



With Categories



With Categories

```
1 > lm(Petal.Length ~ Sepal.Length + Species, iris)
```

```
2
```

```
3 Coefficients:
```

```
4      (Intercept)      Sepal.Length      versicolor      virginica
5      -1.702         0.632         2.210         3.090
```

