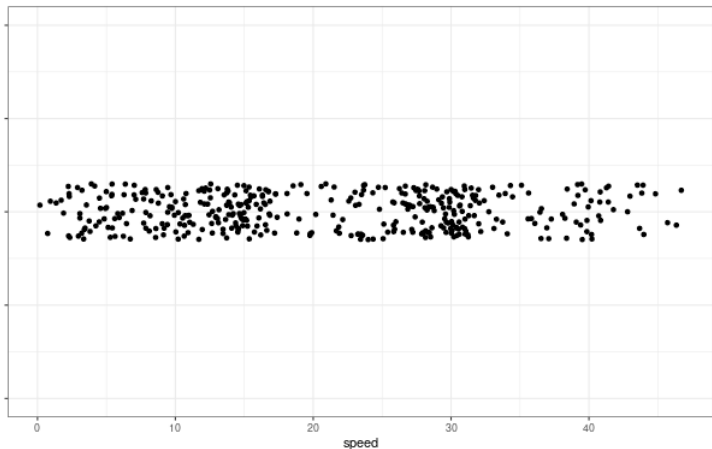# ANOVA Part 2

Grinnell College

April 26, 2024
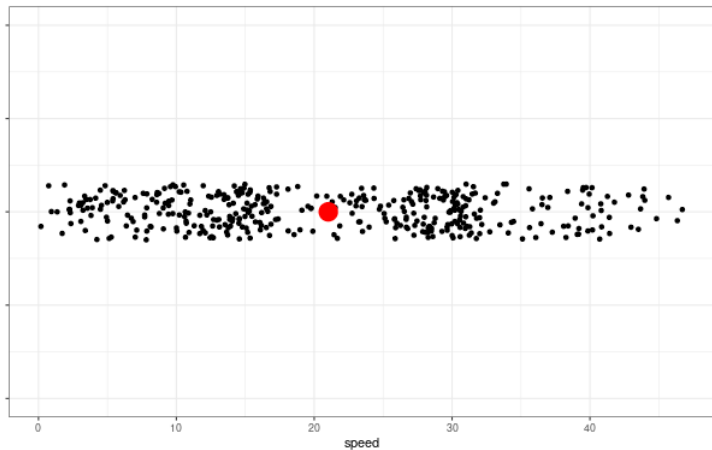
Review

- What are we doing
- What's the point
- Maybe let's have fewer equations today

Speed of all dogs
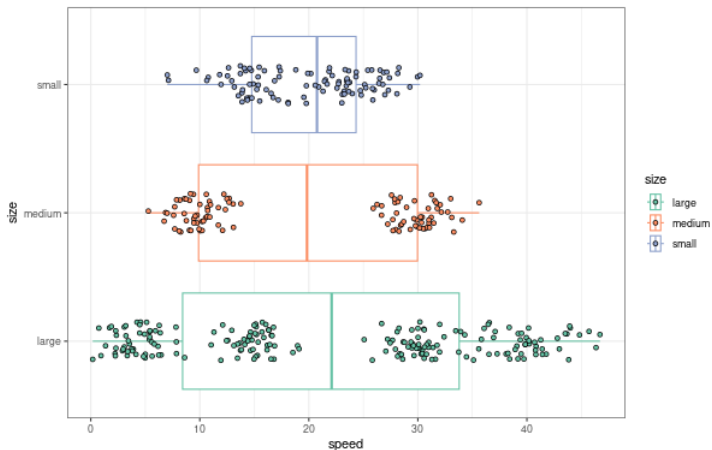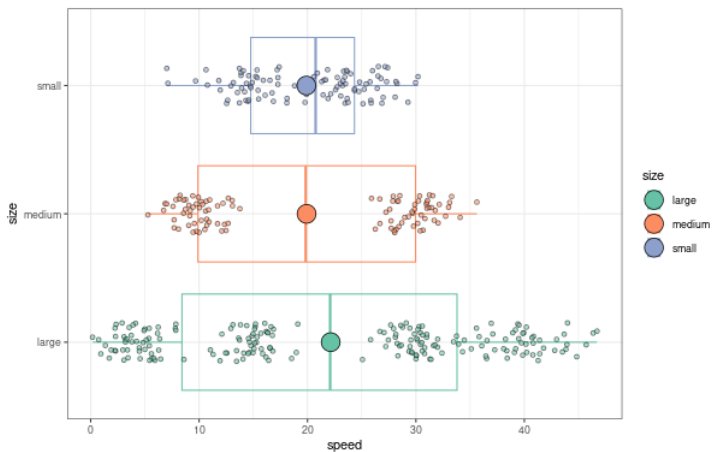
Speed of all dogs

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	399	52398	131.32		

Speed of Size dogs

Speed of Size dogs

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
size	2	498	248.96	1.90	0.1503
Residuals	397	51900	130.73		

Speed of Color dogs
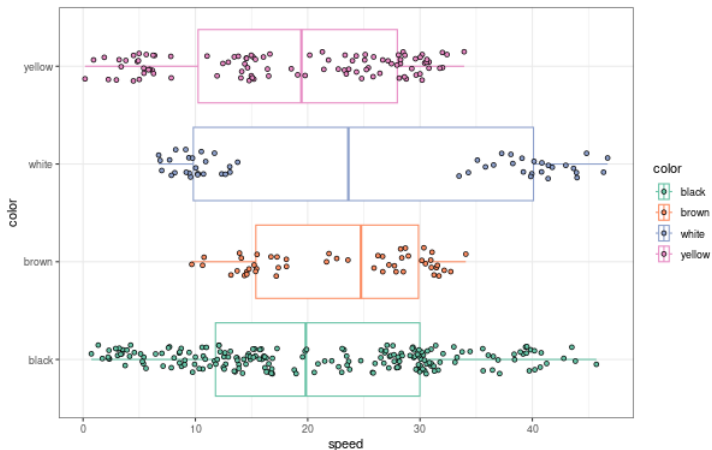
Speed of Color dogs
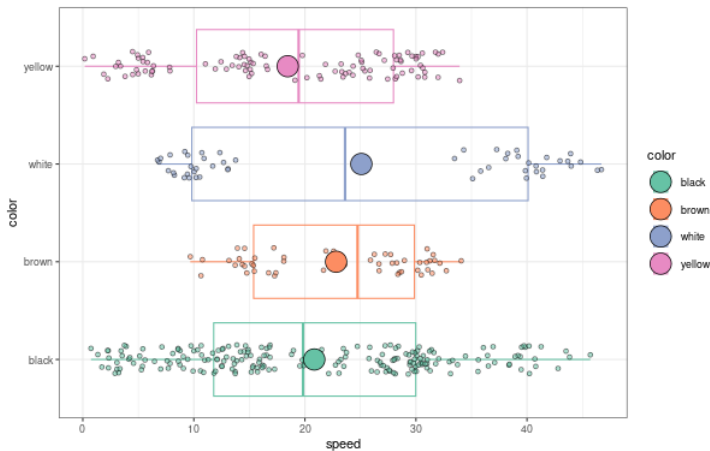
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
color	3	1652	550.55	4.30	0.0053
Residuals	396	50746	128.15		

Speed of Breed dogs
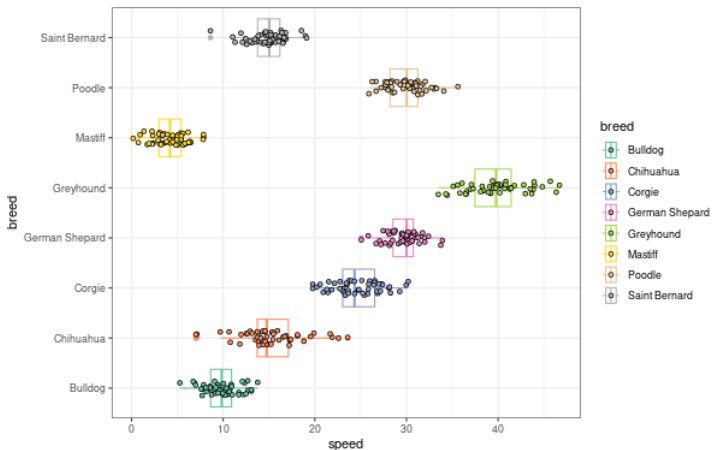
Speed of Breed dogs
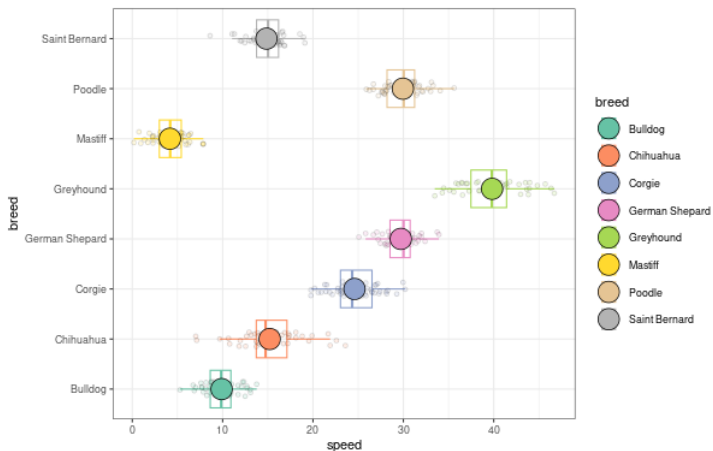
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
breed	7	50066	7152.25	1202.16	0.0000
Residuals	392	2332	5.95		

To t or Not to t

Recall that for ANOVA we are testing the null hypothesis that *all* of our means our equal
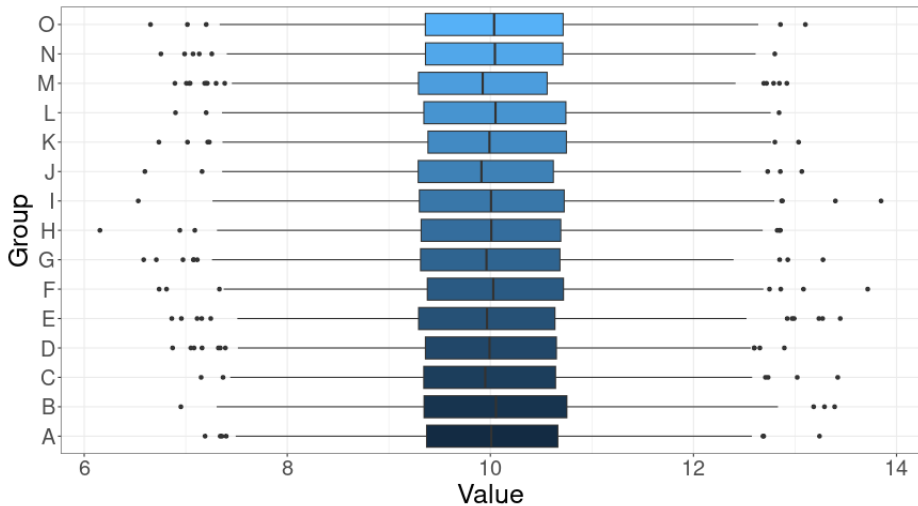
$$H_0 : \mu_A = \mu_B = \mu_C$$

Why not instead just stick with our t-test, doing
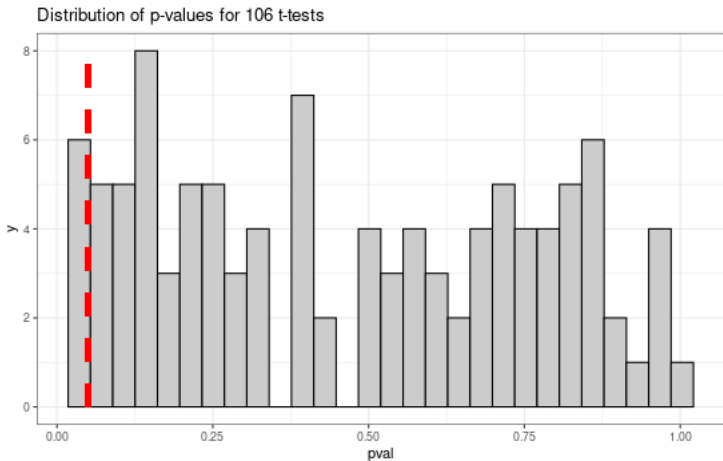
$$H_0 : \mu_A = \mu_B, \ \mu_A = \mu_C, \ \text{and} \ \mu_B = \mu_C$$

Multiple tests

15 groups, all generated with the same mean value:

- ▶ 106 pair-wise tests
- ▶ 6 with p-value < 0.05

Distribution of p-values for 106 t-tests

Multiple tests
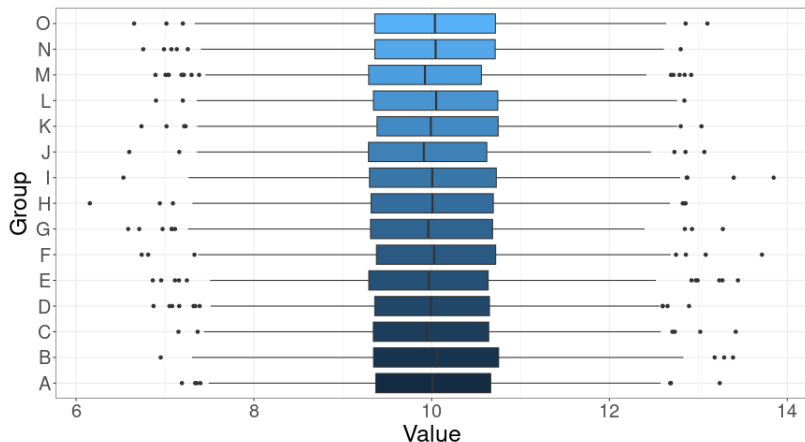
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	14	15.40	1.10	1.10	0.3504
Residuals	14985	14964.85	1.00		

Post-hoc Tests

ANOVA only tells us *that* a difference exists, not where it is or to what degree

If our ANOVA test is such that we rejet the null hypothesis, we can use *post-hoc* testing via the **Tukey Range Test** or the **Tukey Honest Significant Difference Test** to identify any statistically significant pair-wise differences
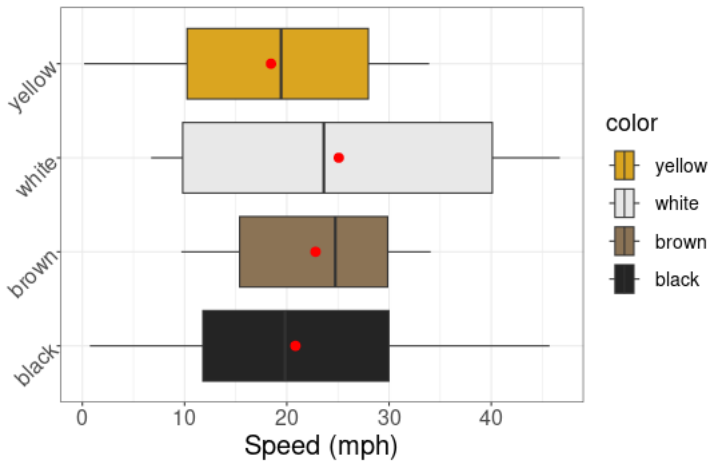
|	Df	Sum Sq	Mean Sq	F value	Pr(>F)
color	3	1651.64	550.55	4.30	0.0053
Residuals	396	50746.34	128.15		

```
1 > aov(speed ˜ color, dogs) %>% TukeyHSD()
2   Tukey multiple comparisons of means
3     95% family-wise confidence level
4
5 Fit: aov(formula = speed ˜ color, data = dogs)
6
7 $color
8                  diff       lwr       upr    p adj
9 brown-black    1.9612  -2.65664   6.57906  0.69237
10 white-black    4.2360  -0.38182   8.85388  0.08529
11 yellow-black  -2.3968  -5.97373   1.18021  0.31012
12 white-brown    2.2748  -3.56635   8.11599  0.74672
13 yellow-brown  -4.3580  -9.41657   0.70063  0.11889
14 yellow-white  -6.6328 -11.69139  -1.57418  0.00437
```

```
1 > aov(speed ~ color, dogs) %>% TukeyHSD() %>% plot()
```


95% family-wise confidence level

Differences in mean levels of color

Review
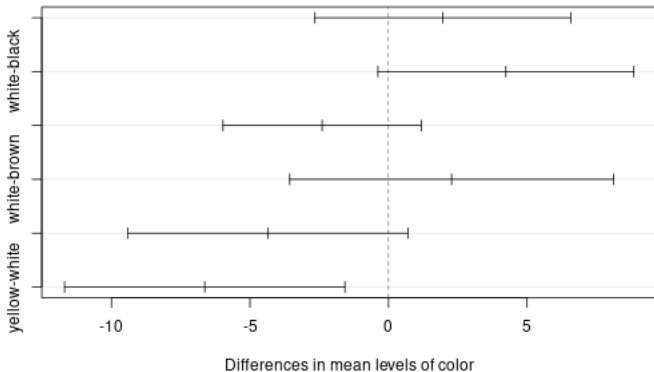
- ▶ ANOVA allows us to test equality of many means
  - ▶ By comparing ratio of between-group and within-group means
- ▶ Ameliorates problem of multiple testing
- ▶ *Post-hoc* testing can be done to determine which groups are different
- ▶ Tukey Honest Statistical Difference (TukeyHSD)