ANOVA:
One-way ANOVAs (practice)

Research Methods for Human Inquiry
Andrew Perfors

This lecture:

- How to do the ANOVA in R
- Which groups actually differ from one another

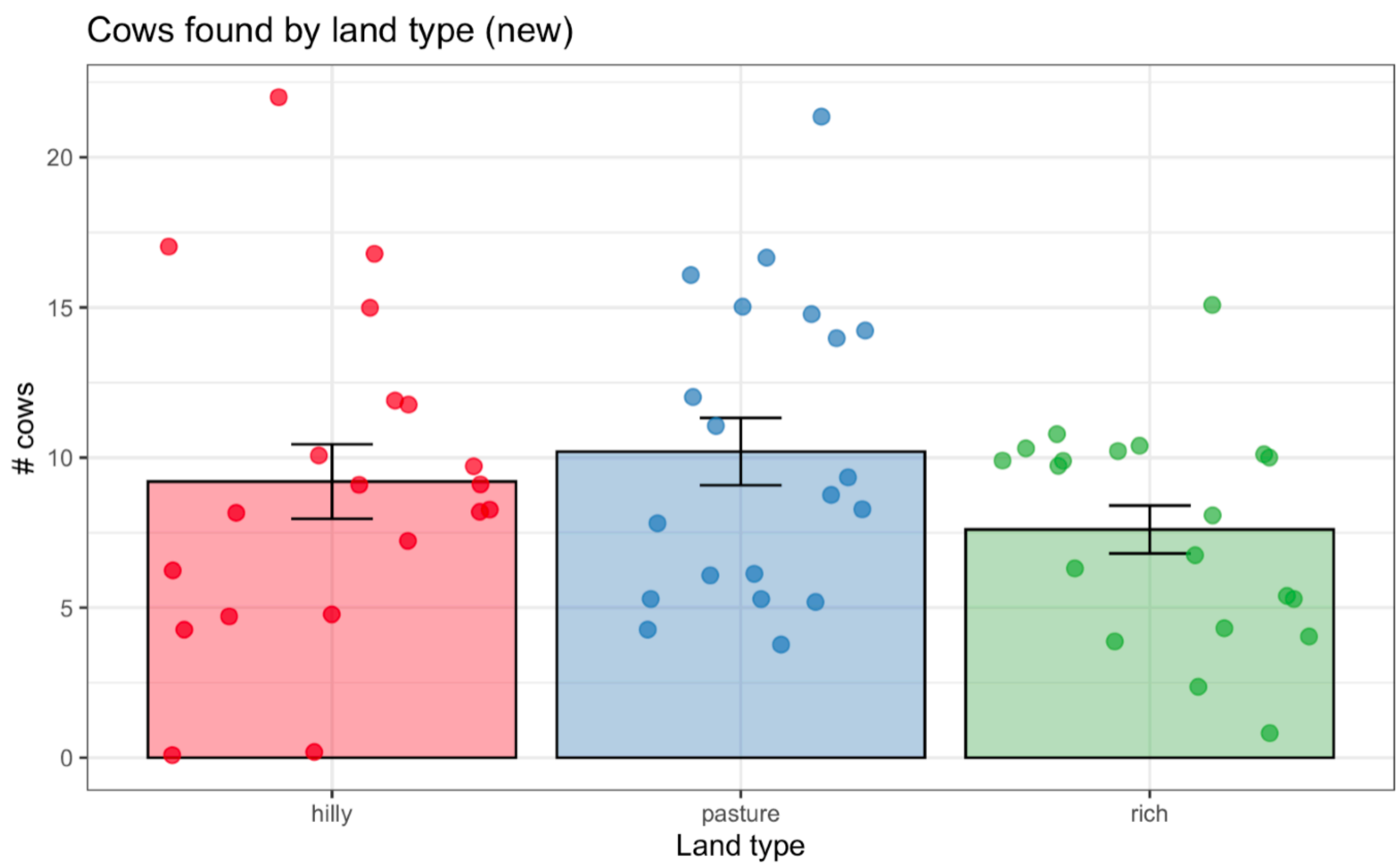
Reminder

- One data frame, `d_new`
- Five variables:
  - `plot`... code uniquely identifying each plot of land
  - `type`... type of land: pasture, rich, or hilly
  - `cows, berries, corns`... units of each food

```
> d_new
      plot    type cows berries corn
1 9Jx21zaa pasture    4    58.7 28.2
2 Qp72PepB pasture   14    57.1 28.3
3 5YIxvYbz pasture   11    31.6 27.9
4 OnZuUW5M pasture    9    34.8 19.9
5 rLMa3j90 pasture    4    52.8 23.0
6 k3Hb2fUa pasture   16    57.9 35.8
```

Reminder

- Are there significantly more cows on pasture right now, as you'd expect if people were allocating land sensibly?



d_new

Analysis of variance

- ANOVA is performed in stages
  - 1. `aov()` calculates the SS values etc
  - 2. `summary()` runs the hypothesis tests
  - 3. other functions to pull out other things of interest

- The `aov()` function
  - This is the main "workhorse" function
  - It creates an "`aov`" object (i.e. variable), which contains lots of quantities of interest relating to ANOVA
  - Let's see how this works in practice...

Using the `aov()` function

```
> cows1waynewModel <- aov( cows ~ type, data = d_new )
```

The model formula:
- `cows` is the outcome variable (DV)
- `type` is the predictor (IV)

Store the output as a variable called `cows1waynewModel`

This tells R to look for the variables inside the `d_new` tibble

What precisely is this thing?

- Our `cows1waynewModel` variable is an "aov object"
  - It's a special kind of variable that stores a whole bunch of information relevant to an ANOVA
  - Printing an aov object will show you only some of the information in it — good because most of it we don’t care about!
  - We use "extractor" functions to pull out the important bits

One way ANOVA

```
> cows1waynewModel <- aov( cows ~ type, data = d_new )
> cows1waynewModel

Call:
   aov(formula = cows ~ type, data = d_new)

Terms:
                 type Residuals
Sum of Squares   68.8    1311.2
Deg. of Freedom     2        57
```

SS_b $G\text{-}1$ SS_w $N\text{-}G$

R has calculated the quantities needed for the F-test

$$MS_w = \frac{SS_w}{N-G} \qquad MS_b = \frac{SS_b}{G-1}$$

$$F = \frac{MS_b}{MS_w}$$

```
> F = (68.8/2) / (1311.2/57)
> F
[1] 1.495424
```

Doing the hypothesis test

Use the function `summary()`

```
> cows1waynewModel <- aov( cows ~ type, data = d_new )
> summary(cows1waynewModel)

            Df Sum Sq Mean Sq F value Pr(>F)
type         2   68.8    34.4   1.495  0.233
Residuals   57 1311.2    23.0
```

The ANOVA table

Each row corresponds to one source of variation: "type" = between groups (i.e., between land types); "residuals" = within groups

F-statistic is ratio of MS values

R also calculates the p-values

```
            Df Sum Sq Mean Sq F value Pr(>F)
type         2   68.8    34.4   1.495  0.233
Residuals   57 1311.2    23.0
```

Degrees of freedom... G=3 groups means that we have G-1 = 2 between-groups, N=60 observations means that we have N-G = 57 within groups

Sum of squares (SS_b and SS_w)

Divide SS by corresponding df to get MS

# of stars gives different levels of significance

A simple measure of effect size

- The η^2 (eta-squared) statistic is conventional
- It's calculated by dividing SS_b by SS_{tot}

$$\eta^2 = \frac{\mathrm{SS}_b}{\mathrm{SS}_{tot}}$$

- It's interpreted as the proportion of the total variance attributable to the grouping variable
  - e.g., $\eta^2 = 0.07$ means 7% of the variation in the outcome variable is explained by the groups, which isn't very much.

How to calculate it in R

- Assuming you've already run an ANOVA and stored it as a variable, then it's really easy
  - Use the `etaSquared()` function [`lsr` package].
  - It only has one argument, the aov object itself:

```
> etaSquared(cows1waynewModel)
         eta.sq eta.sq.part
type 0.04985507  0.04985507
```

This is the part of the output that matters for our purposes. Partial eta-squared is identical to eta-squared for a one-way ANOVA, so we can ignore the second column

Adding effect sizes to your write up?

- It's a good idea to add a measure of effect size to the "stat block". So, instead of writing this...

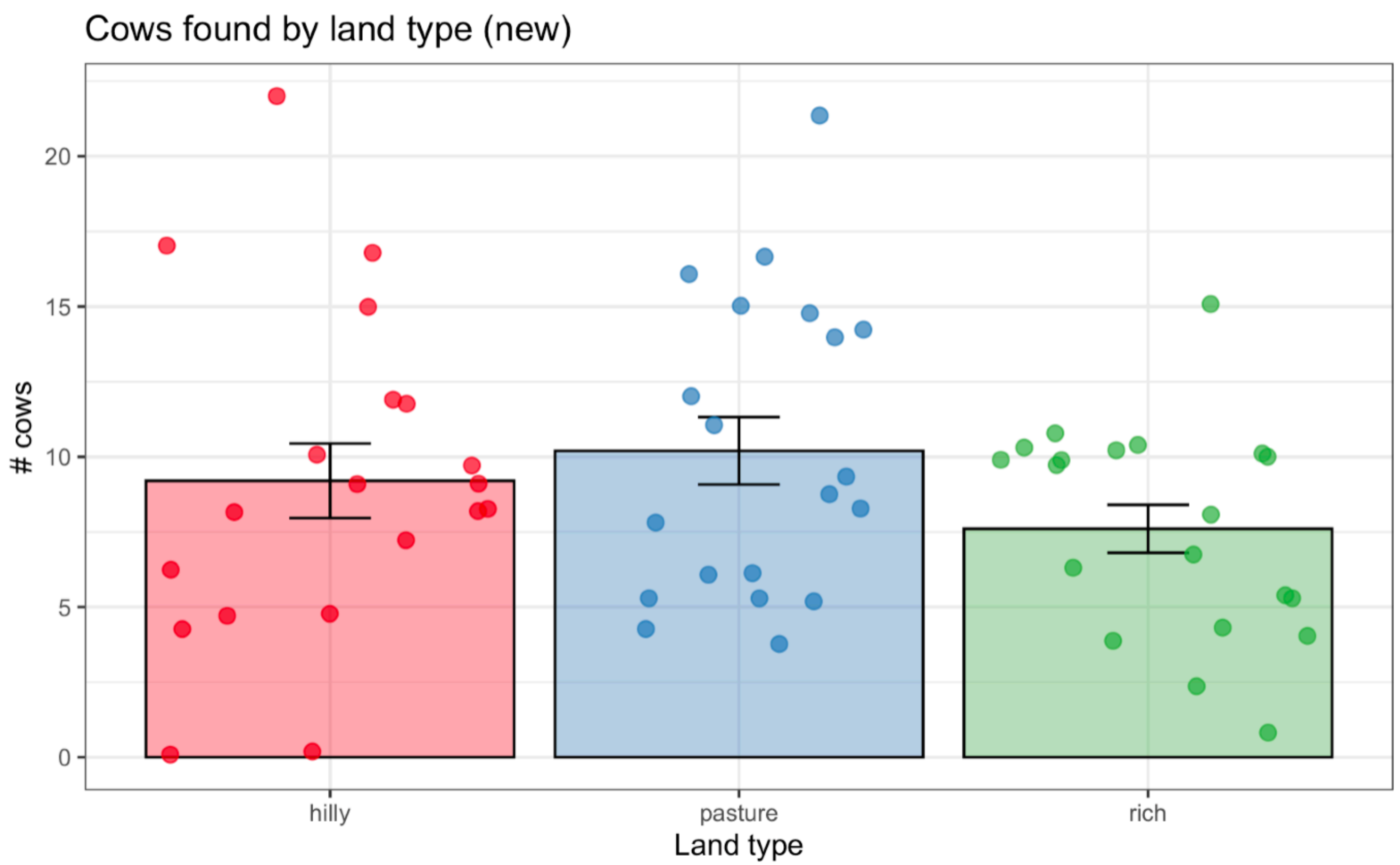
$$F(2,57) = 1.50,\ p = .233$$

- We would write this...

$$F(2,57) = 1.50,\ p = .233,\ \eta^2 = .05$$

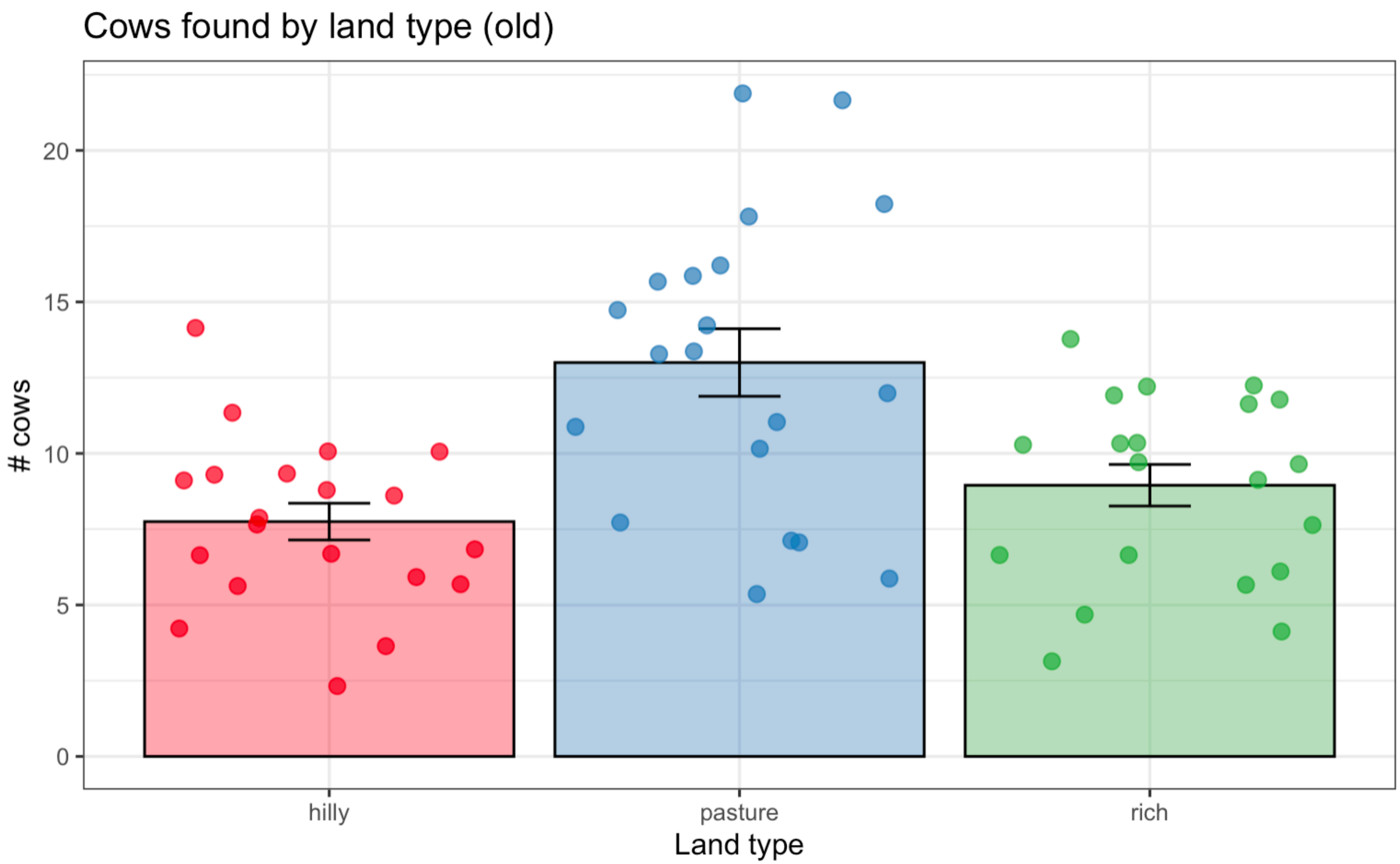
- ... and we might comment that “land type" explains 5% of the variance in the # of cows ranched on that land. It’s a good idea to include effect size and its interpretation!

What does this mean?



There’s no significant difference between land types in # of cows right now, which suggests that maybe they aren’t using the land optimally

Can do the exact same analysis on the old data…



d_old

Can do the exact same analysis on the old data...

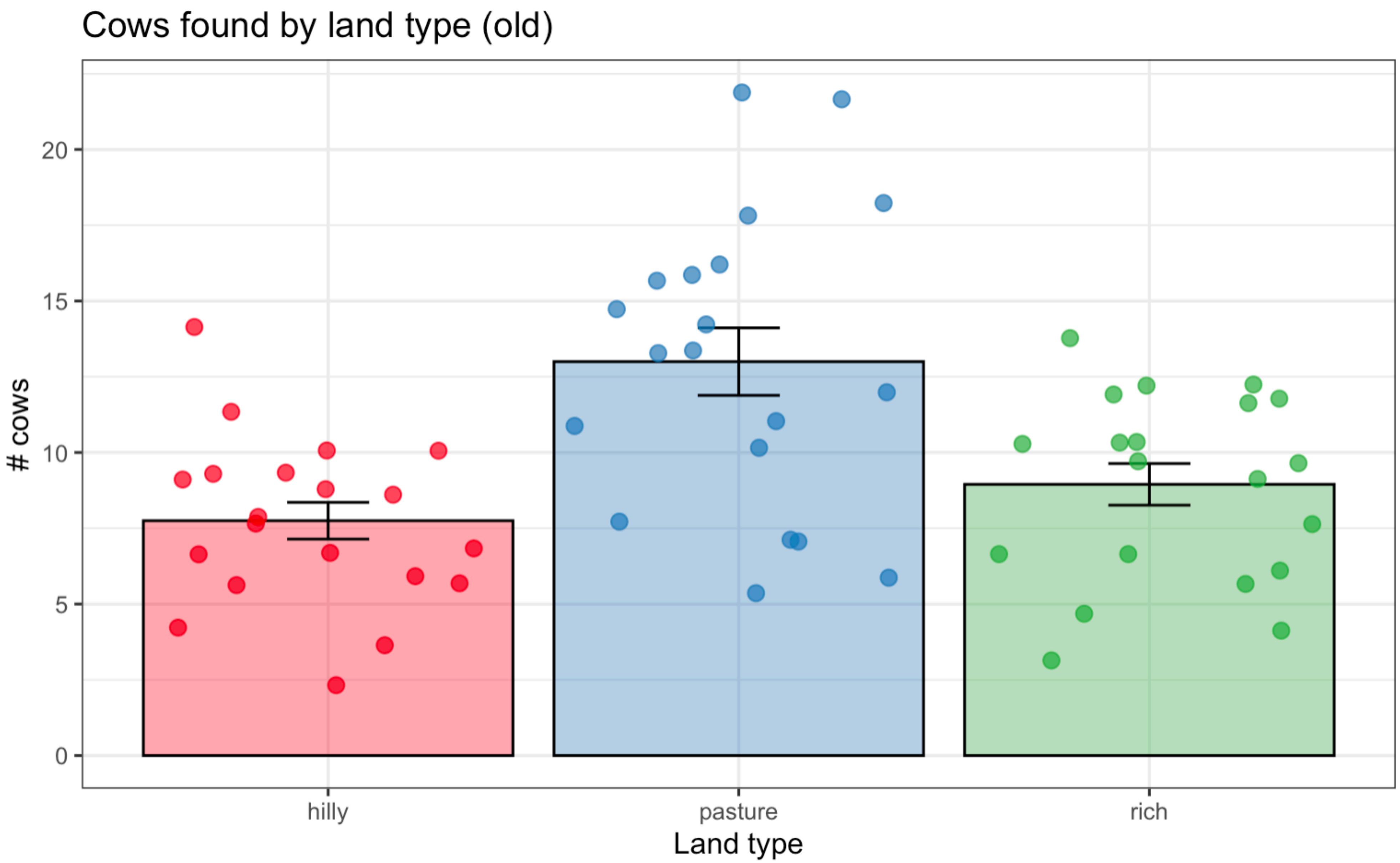
```
> cows1wayoldModel <- aov( cows ~ type, data = d_old )
> summary(cows1wayoldModel)

            Df Sum Sq Mean Sq F value   Pr(>F)
type         2  302.7  151.35   10.91 9.73e-05 ***
Residuals   57  790.7   13.87
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1

> etaSquared(cows1wayoldModel)

        eta.sq eta.sq.part
type 0.2768429   0.2768429
```

This is significant



Analysis suggests that 15 years ago, land type explained 27% of the variance in # of cows - perhaps they used land more optimally back then

Exercises are in w8day1exercises.Rmd