

T-tests:

Paired samples

Research Methods for Human Inquiry
Andrew Perfors

Today's story...

The economy doesn't explain everything, so what is going on?



Maybe crop yields are going down - I've heard of things like over-farming, maybe that's happening?

Today's story...

The economy doesn't explain everything, so what is going on?

For each bit of cropland we could compare its yield this year compared to five years ago

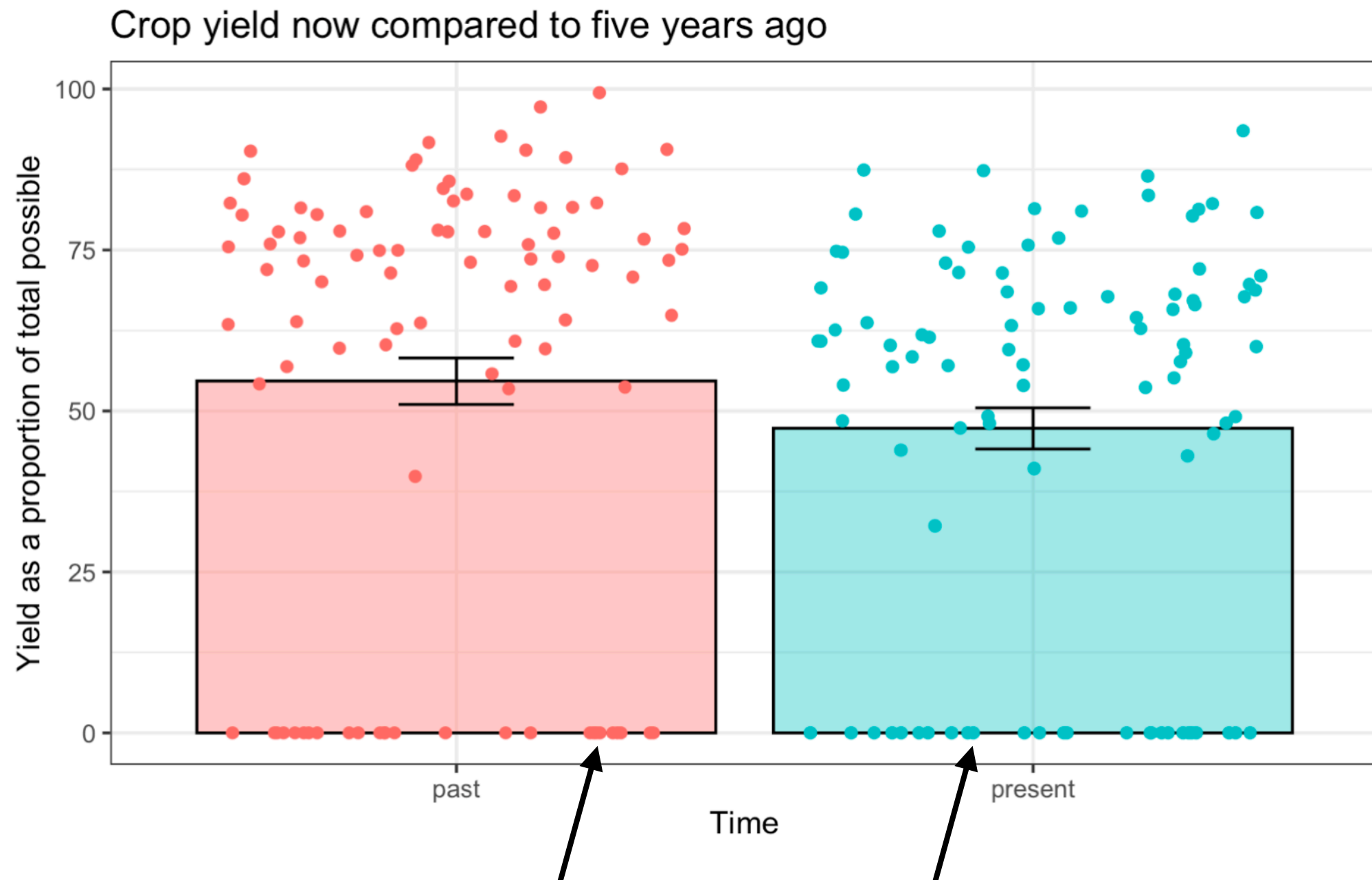


Comparing two means
within a repeated measures design,
using a **paired samples t-test**

Suppose you have repeated measurements?

	5 years ago	Now
Location1	63.7	47.3
Location2	62.9	53.6
Location3	90.5	67.7
Location4	71.4	63.3
Location5	60.8	57.9

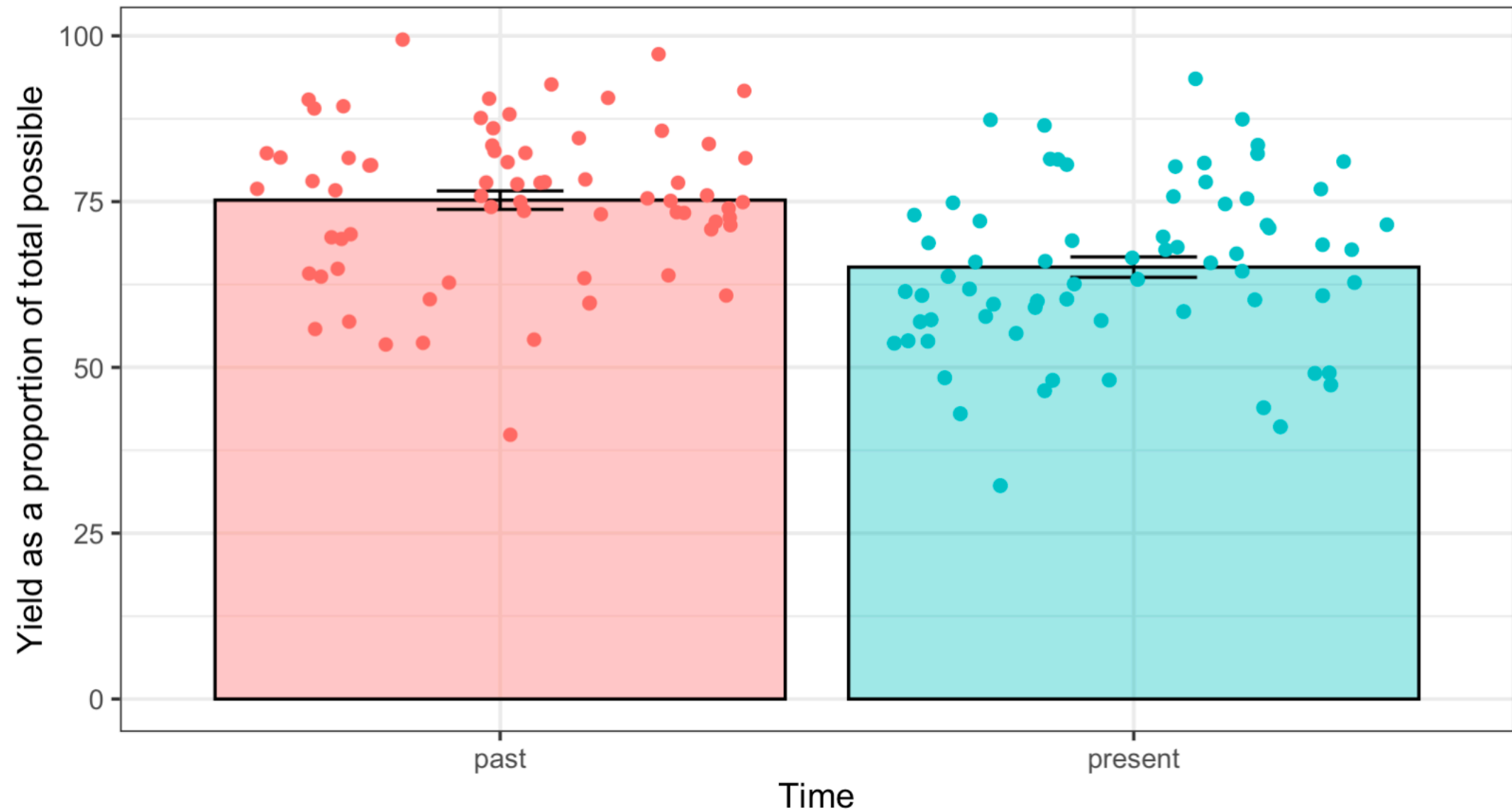
As always, let's graph them...



That's odd — probably don't want to include these,
since no crops are being grown on this land at all

As always, let's graph them...

Crop yield now compared to five years ago
(Completely nonproductive land removed)



But this is still not *quite* what we're interested in!

What are we really interested in here?

- We don't want to compare groups, we want to look at changes within individuals!
 - What we actually want to know is whether each location had its crop yield go up, down, etc
 - That is, we're interested in the difference scores...

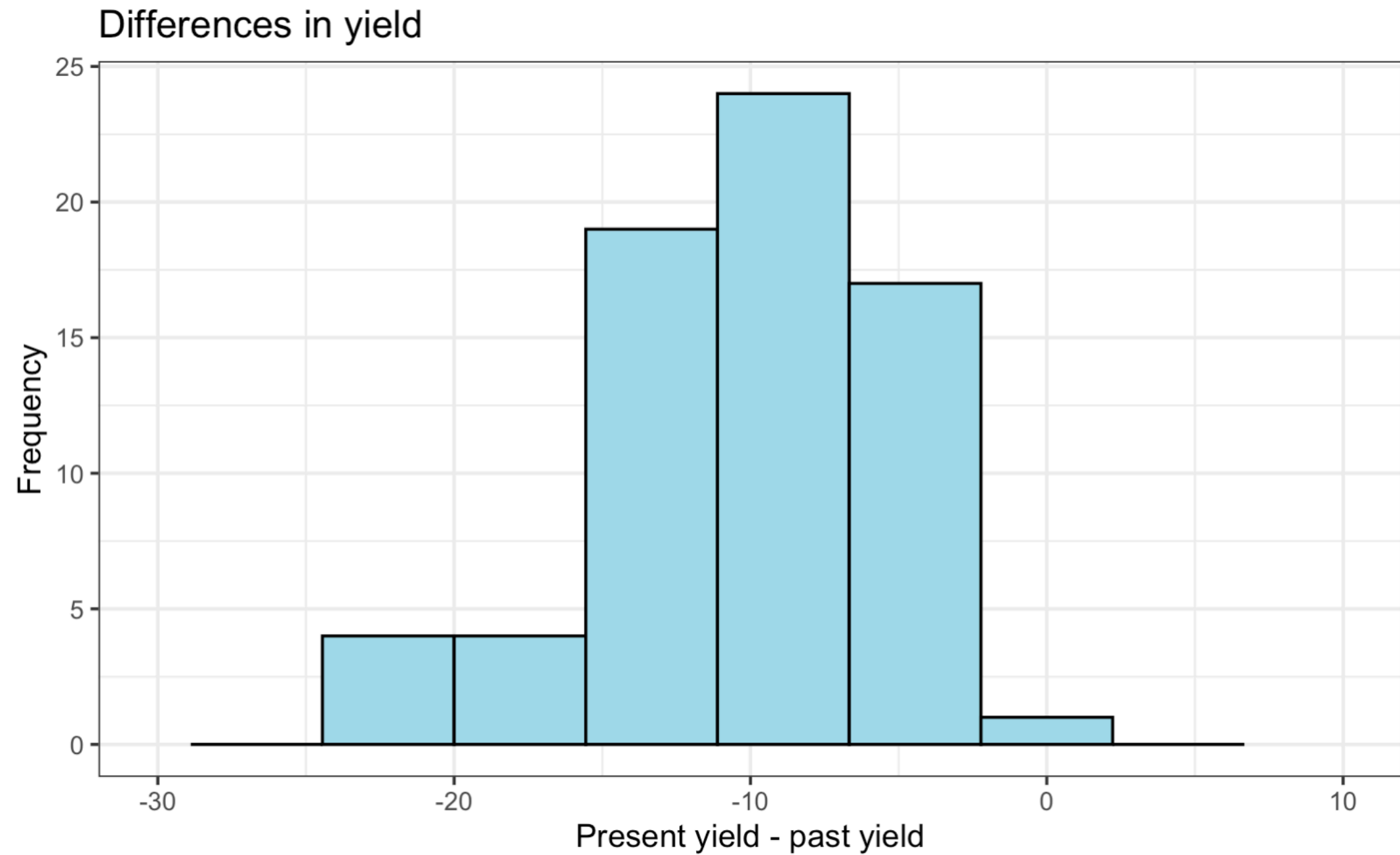
Paired samples t-test

- How the paired samples t-test works
 - Run a **one sample** t-test on the difference scores
 - Null hypothesis: mean difference is 0
 - Alternative hypothesis: mean difference is not 0

	5 years ago	Now	Change
Location1	63.7	47.3	-16.4
Location2	62.9	53.6	-9.3
Location3	90.5	67.7	-22.8
Location4	71.4	63.3	-8.1
Location5	60.8	57.9	-2.9

Is this
different
from zero?

Let's look at *this* data



Paired samples t-test

- How the paired samples t-test works
 - Run a **one sample** t-test on the difference scores
 - Null hypothesis: mean difference is 0
 - Alternative hypothesis: mean difference is not 0
- There is nothing conceptually interesting here
 - It is exactly the same test as a one-sample t-test
 - So I won't bother repeating myself telling you how it works, because you already know.
 - Instead, let's jump to "doing it in R"...

Our data can be in both wide form and long form

```
> head(dc2)
# A tibble: 6 x 3
  location  past present
  <chr>    <dbl>   <dbl>
1 IOL7lwIV 69.6     56.9
2 AGfZ9hKP 63.5     60.9
3 Vz10ctUb 78.0     71.5
4 RpLC77GF 82.3     75.8
5 Ggynz3I1 81.6     73.0
6 5Nt0dKug 55.8     43.9
```

wide form: different columns
for time 1 and time 2

```
> head(dc_long2)
# A tibble: 6 x 3
  location time      yield
  <chr>    <chr>    <dbl>
1 IOL7lwIV past      69.6
2 IOL7lwIV present   56.9
3 AGfZ9hKP past      63.5
4 AGfZ9hKP present   60.9
5 Vz10ctUb past      78.0
6 Vz10ctUb present   71.5
```

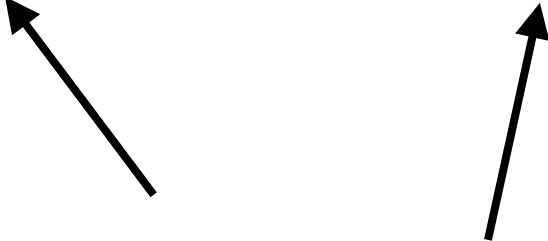
long form: one variable
indicates which time it was

Wide form paired-sample t-test in R

You can use the `t.test()` command

```
> t.test(x=dc2$past, y=dc2$present, paired=TRUE)
```

Two variables, `x` and `y`. Assumes they
each have the same # of rows

Two black arrows originate from the text 'Two variables, x and y. Assumes they each have the same # of rows'. One arrow points diagonally up and to the right towards the parameter 'x=dc2\$past' in the code. The other arrow points diagonally up and to the left towards the parameter 'y=dc2\$present' in the code.

indicates a paired-sample t-test rather
than independent-sample t-test

A black arrow originates from the text 'indicates a paired-sample t-test rather than independent-sample t-test'. It points diagonally up and to the left towards the parameter 'paired=TRUE' in the code.

Wide form paired-sample t-test in R

You can use the `t.test()` command

```
> t.test(x=dc2$past, y=dc2$present, paired=TRUE)
```

Paired t-test

data: dc2\$past and dc2\$present

t = 17.693, df = 68, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

8.961695 11.240102

sample estimates:

mean of the differences

10.1009

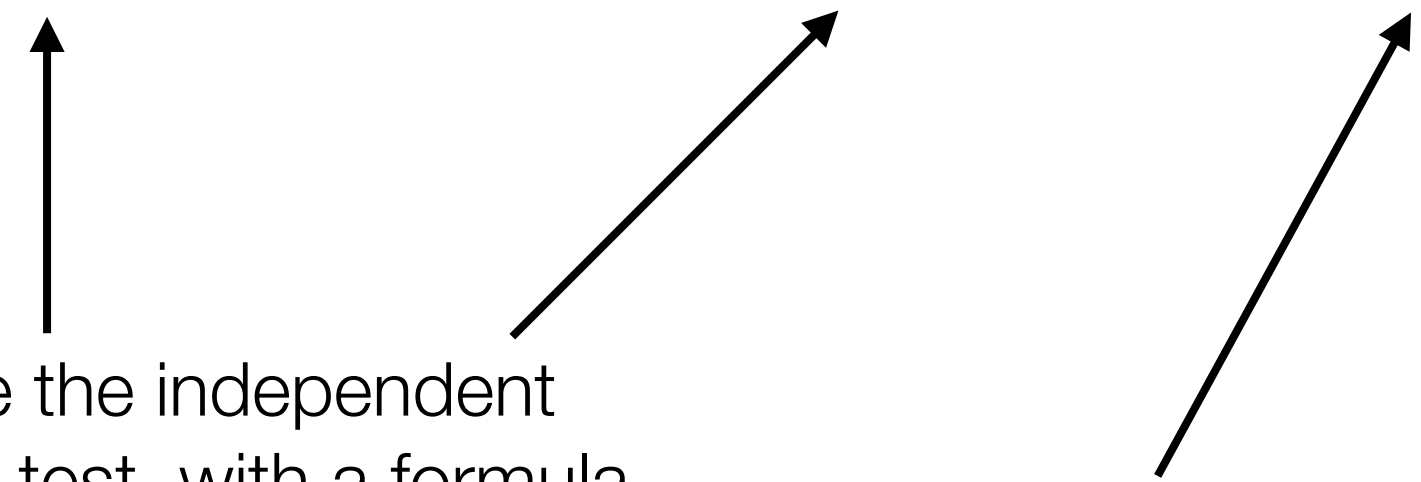
all standard output, except here it shows
it is a paired t-test and is looking at the
difference in means (subtracts `y` from `x`)

and the confidence interval is on the size of the difference score

Long form paired-sample t-test in R

You can use the `t.test()` command

```
> t.test(formula = yield ~ time, data=dc_long2, paired=TRUE)
```



Just like the independent
samples t-test, with a formula
and dataset

indicates a paired-sample t-test rather
than independent-sample t-test

Long form paired-sample t-test in R

You can use the `t.test()` command

```
> t.test(formula = yield ~ time, data=dc_long2, paired=TRUE)
```

Paired t-test

data: yield by time

t = 17.693, df = 68, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

8.961695 11.240102

sample estimates:

mean of the differences

10.1009

Exercises are in `w7day2exercises.Rmd`