From Longer and Wider, We Stand on Guard for Thee

Ken Butler Department of Computer and Mathematical Sciences
UTSC (Scarborough)

Packages

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
## -- Conflicts ------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Pig feed

20 pigs are randomly assigned to one of four pig feeds, and the weight gain of each pig is measured:

pig	feed1	feed2	feed3	feed4
1	60.8	68.7	92.6	87.9
2	57.0	67.7	92.1	84.2
3	65.0	74.0	90.2	83.1
4	58.6	66.3	96.5	85.7
5	61.7	69.8	99.1	90.3

Say we want graphs of weight gain for each feed.

Read in:

```
pigs <- read_table("pigs1.txt")
```

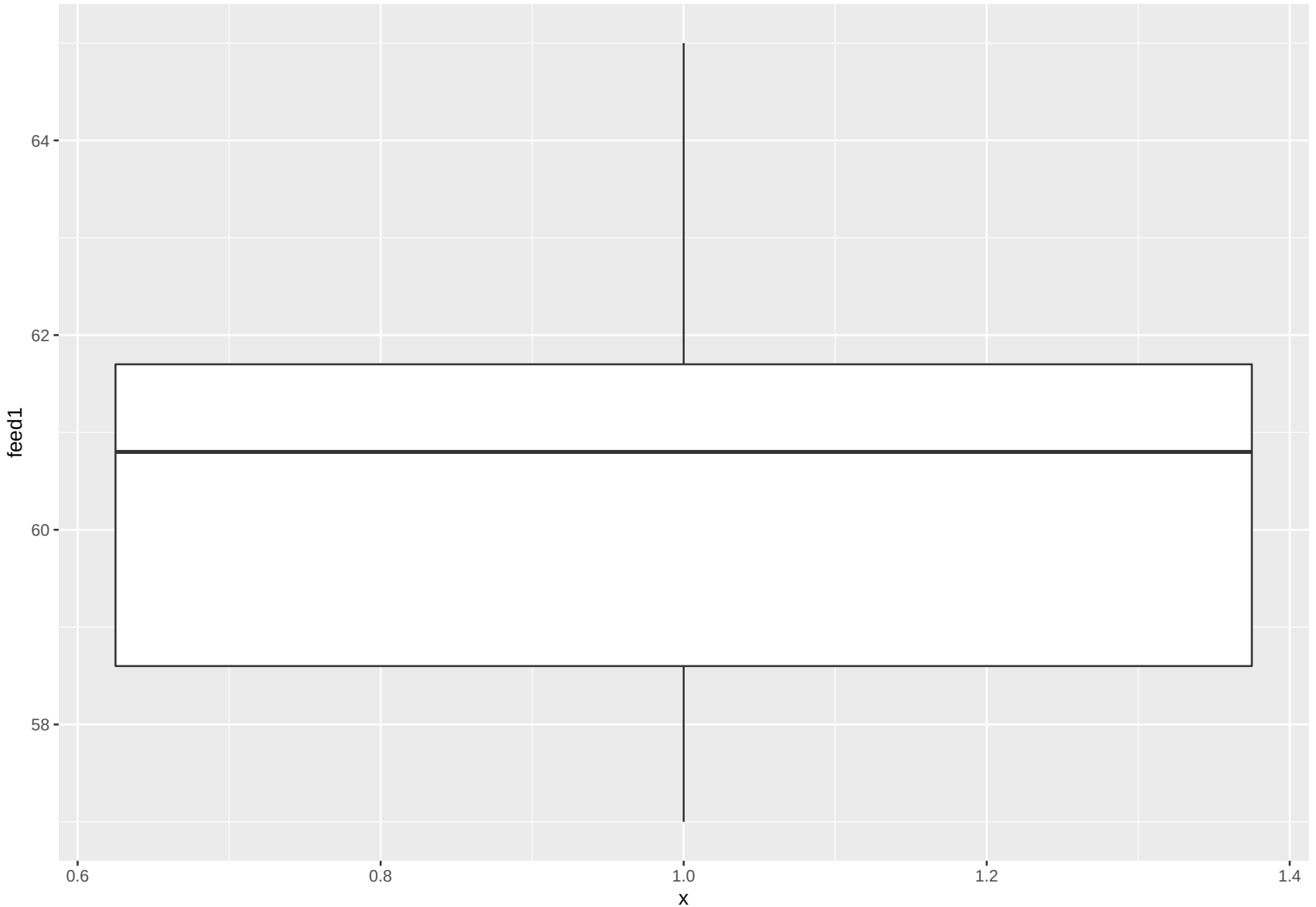
```
## Parsed with column specification:
## cols(
##   pig = col_double(),
##   feed1 = col_double(),
##   feed2 = col_double(),
##   feed3 = col_double(),
##   feed4 = col_double()
## )
```

```
pigs
```

```
## # A tibble: 5 x 5
##     pig feed1 feed2 feed3 feed4
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1  60.8  68.7  92.6  87.9
## 2     2  57    67.7  92.1  84.2
## 3     3  65    74    90.2  83.1
```

and then we have to do this 4 times...

```
ggplot(pigs, aes(x=1, y=feed1)) + geom_boxplot()
```



The problem

- The data frame is the *wrong shape*.
- Need all the weight gains in *one* column, with another column saying what feed that weight gain was from
- Make data frame longer.
- Old tools:
  - `reshape`
  - `reshape2`
  - `gather` (from `tidyr`)
- New tool: `pivot_longer`

On the pig feed data

```
pigs
```

```
## # A tibble: 5 x 5
##     pig feed1 feed2 feed3 feed4
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1  60.8  68.7  92.6  87.9
## 2     2  57    67.7  92.1  84.2
## 3     3  65    74    90.2  83.1
## 4     4  58.6  66.3  96.5  85.7
## 5     5  61.7  69.8  99.1  90.3
```

```
pigs %>% pivot_longer(-pig, names_to="feed",
                      values_to="weight") -> pigs_longer
```

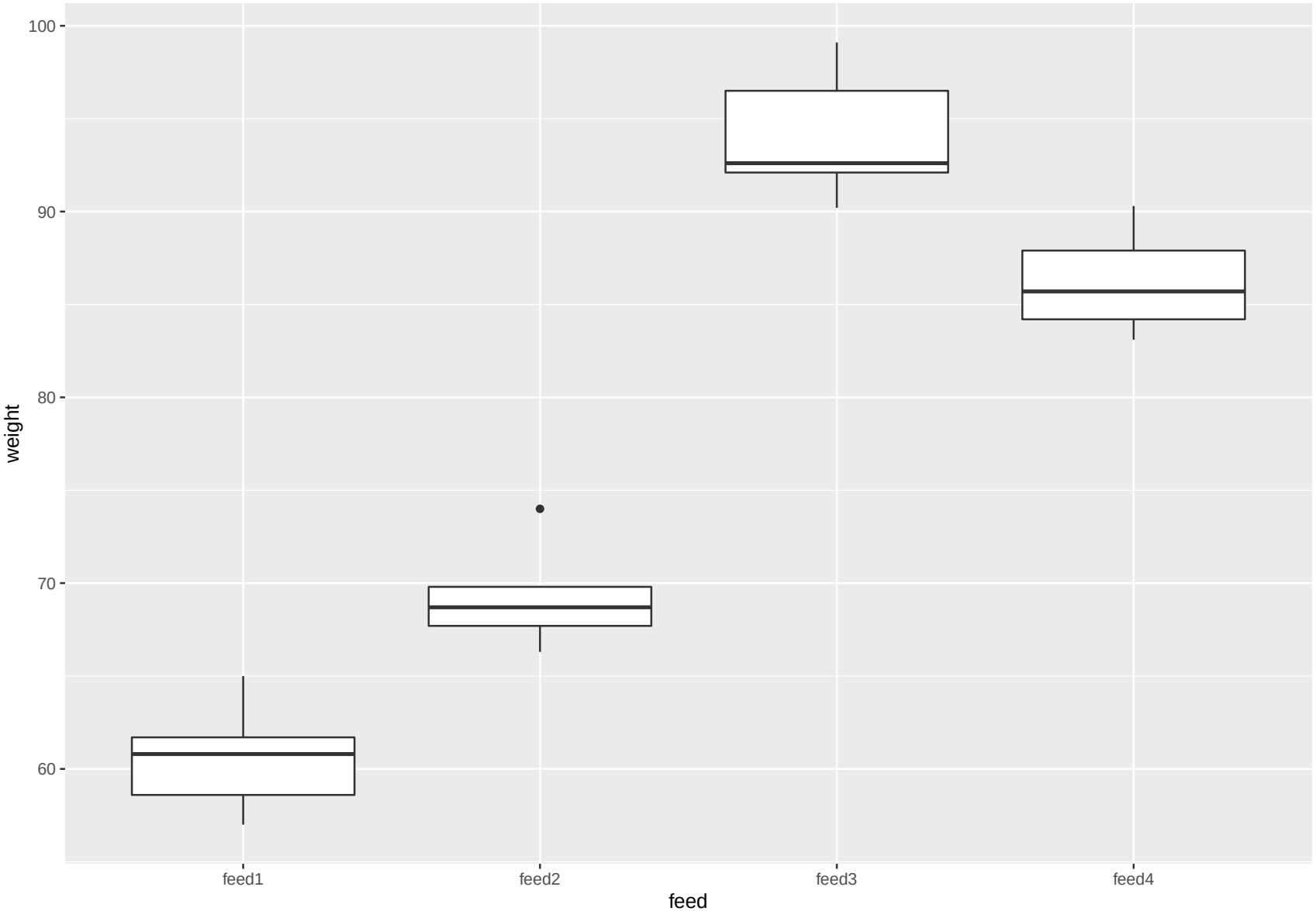
The results

```
pigs_longer
```

```
## # A tibble: 20 x 3
##      pig feed  weight
##    <dbl> <chr>  <dbl>
##  1     1 feed1   60.8
##  2     1 feed2   68.7
##  3     1 feed3   92.6
##  4     1 feed4   87.9
##  5     2 feed1   57
##  6     2 feed2   67.7
##  7     2 feed3   92.1
##  8     2 feed4   84.2
##  9     3 feed1   65
## 10     3 feed2   74
## 11     3 feed3   90.2
## 12     3 feed4   83.1
## 13     4 feed1   58.6
## 14     4 feed2   66.3
```

Now we can make all 4 graphs at once

```
ggplot(pigs_longer, aes(x=feed, y=weight)) + geom_boxplot()
```



another example

this one:

Species	Disease present		Disease absent	
	Location X	Location Y	Location X	Location Y
A	44	12	38	10
B	28	22	20	18

one where I use .value

summarize with list, eg the exam one reformat the data so I have to process it first