

Session 2

Lunch R workshop 2018.03.14

Max Lindmark & Philip Jacobson

lunchR

NEVER HAVE I FELT SO
CLOSE TO ANOTHER SOUL
AND YET SO HELPLESSLY ALONE
AS WHEN I GOOGLE AN ERROR
AND THERE'S ONE RESULT
A THREAD BY SOMEONE
WITH THE SAME PROBLEM
AND NO ANSWER
LAST POSTED TO IN 2003

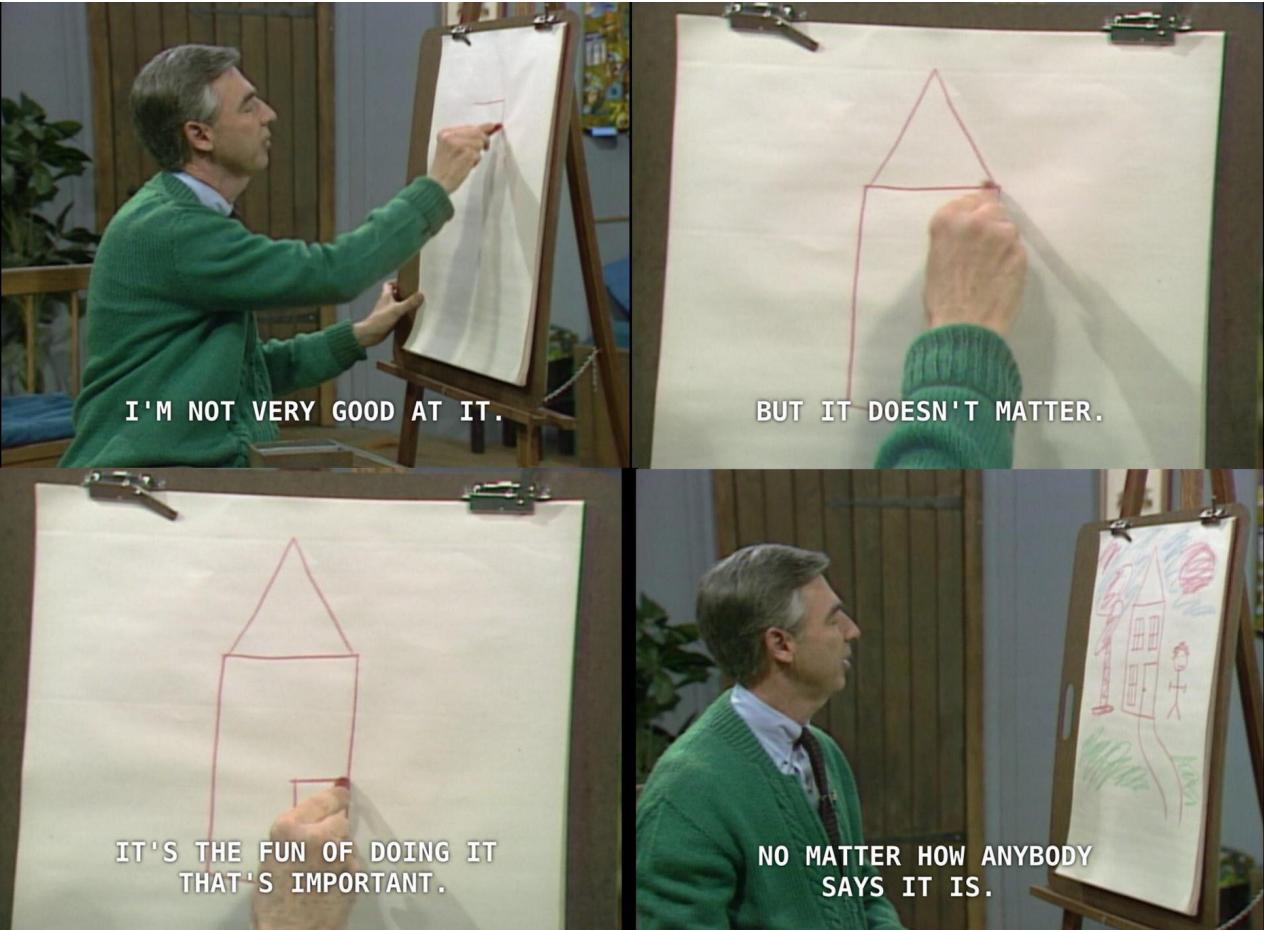
WHO WERE YOU,
DENVERCODER?!

WHAT DID YOU SEE?!



Sorry to disappoint you again
but we are not experts...

LunchR



... we only enjoy doing it!
(and we hope to learn
from you too!)

Sessions

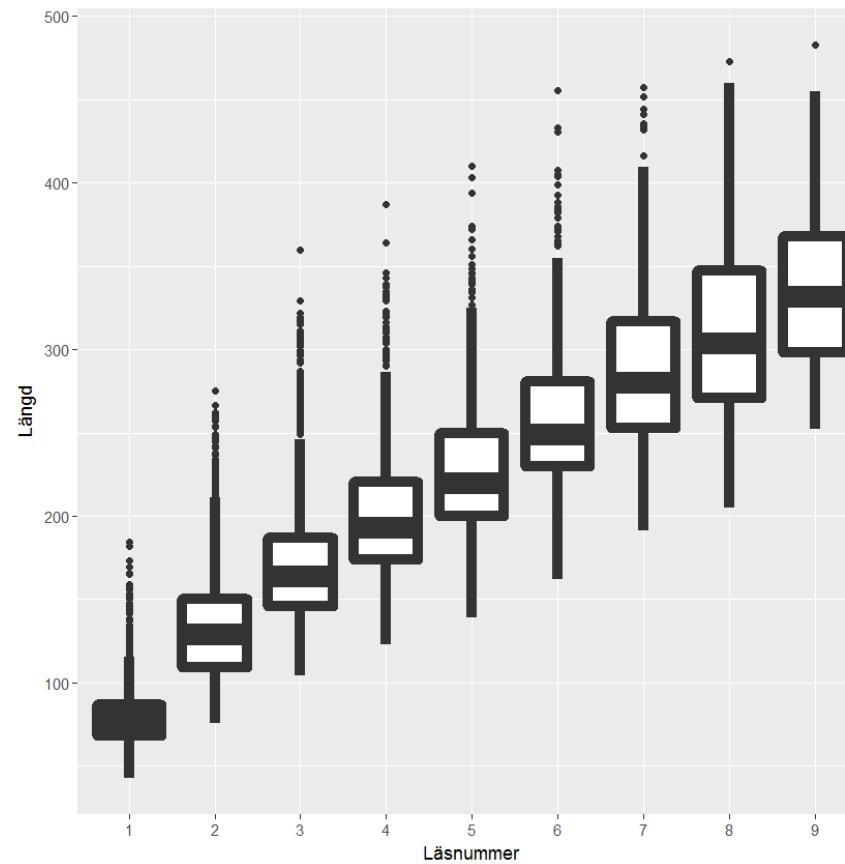
1. Intro to why R, data formats and plotting styles in R
2. **Digging deeper into data exploration in R using base R and ggplot; focus on different graph styles**
3. Tips and tricks in R – format your data for easy plotting
4. Creating publish-quality figures

Efficient data visualization

- Plot types!
 - Scatter plots, boxplots, density plots, histogram, multi-panel plots, ridge-plots
 - ggplot vs base - again
- Introducing the pipe! `%>%`
- Surprise

Why data visualization?

- Explore patterns in your data



	Lokal	Fångstår	Födelseår	ID	Ålder	Redskap	Läsnr	Längd
1	BT	2004	2001	2004577	3	9	1	101.39
2	BT	2004	2001	2004577	3	9	2	171.34
3	BT	2004	2001	2004577	3	9	3	218.17
4	BT	2004	2000	2004573	4	9	1	87.25
5	BT	2004	2000	2004573	4	9	2	125.94
6	BT	2004	2000	2004573	4	9	3	184.02
7	BT	2004	2000	2004573	4	9	4	239.56
8	BT	2004	2003	2004572	1	9	1	133.48
9	BT	2004	2000	2004571	4	9	1	111.25
10	BT	2004	2000	2004571	4	9	2	153.65
11	BT	2004	2000	2004571	4	9	3	213.65
12	BT	2004	2000	2004571	4	9	4	269.23
13	BT	2004	2003	2004569	1	9	1	145.19
14	BT	2004	2000	2004568	4	9	1	98.58
15	BT	2004	2000	2004568	4	9	2	142.92
16	BT	2004	2000	2004568	4	9	3	230.47
17	BT	2004	2000	2004568	4	9	4	299.82
18	BT	2004	2001	2004567	3	9	1	103.71
19	BT	2004	2001	2004567	3	9	2	168.61
20	BT	2004	2001	2004567	3	9	3	224.91
21	BT	2004	1999	2004566	5	9	1	86.95
22	BT	2004	1999	2004566	5	9	2	149.03
23	BT	2004	1999	2004566	5	9	3	172.95
24	BT	2004	1999	2004566	5	9	4	231.17
25	BT	2004	1999	2004566	5	9	5	285.29
26	BT	2004	2002	2004565	2	9	1	81.08
27	BT	2004	2002	2004565	2	9	2	230.57
28	BT	2004	2000	2004564	4	9	1	95.70
29	BT	2004	2000	2004564	4	9	2	158.79
30	BT	2004	2000	2004564	4	9	3	218.90
31	BT	2004	2000	2004564	4	9	4	276.11
32	BT	2004	2000	2004563	4	9	1	88.17
33	BT	2004	2000	2004563	4	9	2	132.44
34	BT	2004	2000	2004563	4	9	3	187.87
35	BT	2004	2000	2004563	4	9	4	229.48
36	BT	2004	2003	2004557	1	9	1	169.57
37	BT	2004	2000	2004555	4	9	1	76.01
38	BT	2004	2000	2004555	4	9	2	140.51
39	BT	2004	2000	2004555	4	9	3	208.25

Why data visualization?

- Test if data manipulation is correct

Iris data set (included in R)

- scatter plots
- boxplots
- density plots
- histogram
- multi-panel plots
- ridge-plots

ggplot vs base - again



Introducing the pipe!

```
library(magritter)  
library(dplyr)
```

%<%



"Pipes take the output from one function and feed it to the first argument of the next function."

```
data %>%
  filter(length > 5) %>%
  ggplot(., aes(x = x, y = length)) +
  geom_point()
```

%<%

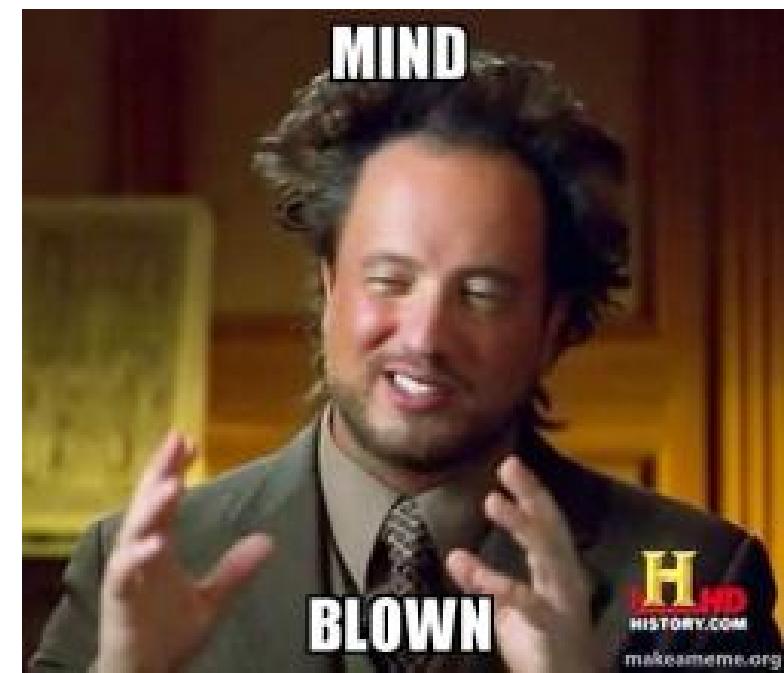


Ceci n'est pas une pipe.

*replace data with a 'dot' when in a pipe!

```
data %>%  
  filter(length > 5) %>%  
  ggplot(., aes(x = x, y = length)) +  
    geom_point()
```

Remember the data only exists with in the pipe



Lets explore some flower morphology!!

