Introduction to R Workshop

Session 3

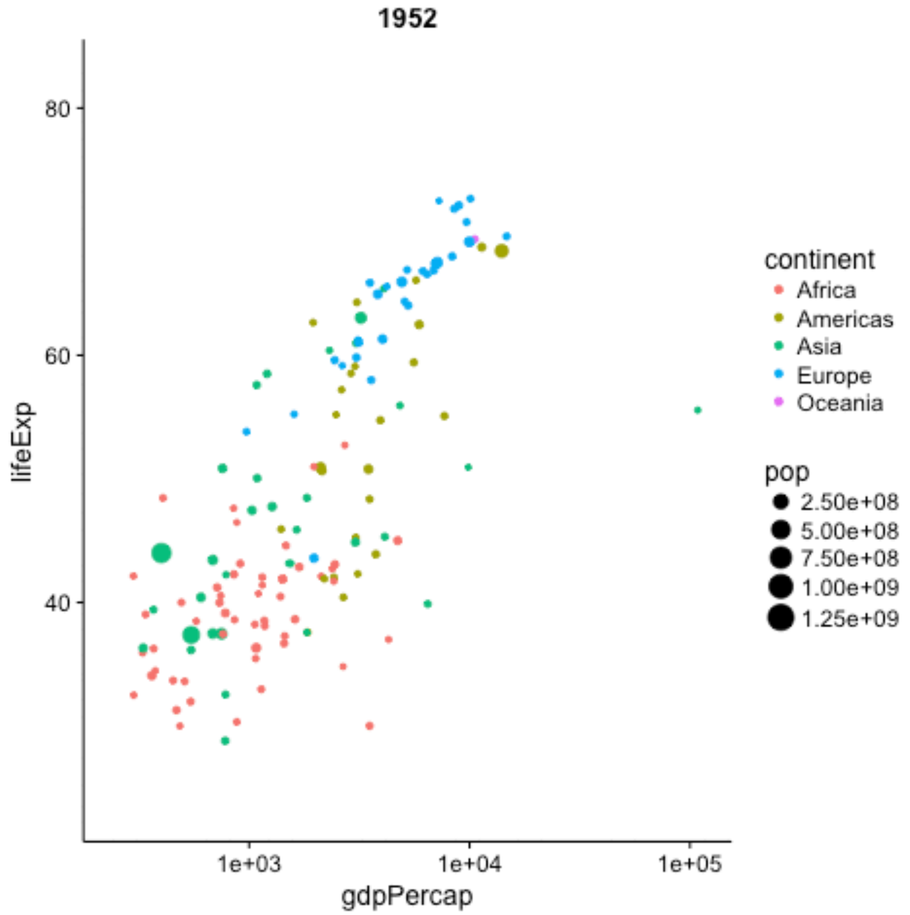
Sean Nguyen

MSU BEST

Broadening Experiences in Scientific Training

Session 3: Goals

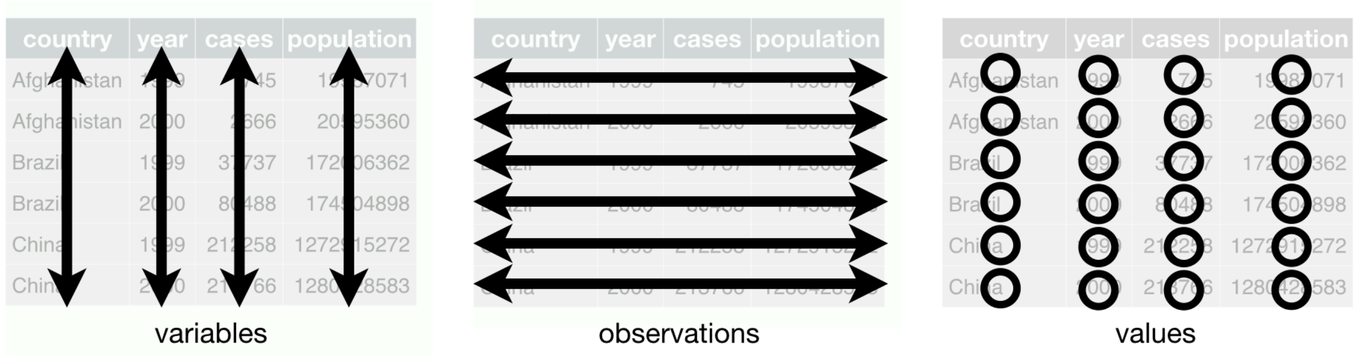
- Learn factors and levels
- Grammar of graphics
- ggplot2
- Saving plots



data types in R:

- Factors - categorical variable Monday , Tuesday, Wednesday
- Numeric - numbers
  - Integers (~2 billion) 1.15, 4.40, 3.80
  - Double (~1.79e308)
- Character - strings of characters “Michigan State University R Workshop”
- Logical - TRUE/FALSE

ggplot2



- Proper formatting
  - tidy data - variables in columns, observations in rows
  - factors - categorical variable
  - integers/numeric - number variable
  - levels - order of categorical variables

ggplot2

- Factors
  - categorical variable

Formula: data$column <- as.factor(data$column)

data$month <- as.factor(data$column)

month (chr)
January
February
March

month (fctr)
January
February
March

ggplot2

- Levels
  - set order of ‘categorical’ factors
  - (R defaults to alphabetical order)

Formula: data$column <- factor(data$column, levels = c(“order”, “that”,”you”, “want”))

data$month <- factor(data$month, levels = c(“January”, “February”, “March”, “April”))

‘April’, ’February’, ’January’, ’March’ → ’January’, ’February’, ’March’, ’April’

ggplot2

General Formula: ggplot(aes(x = ___, y = ____)) +

geom_point()

geom_line()

geom_boxplot()

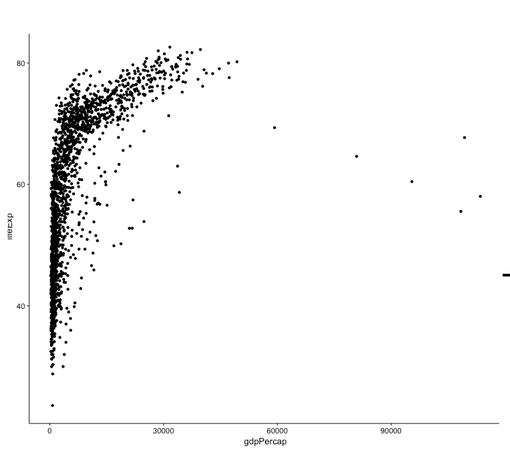
geom_violin()

geom_col()/geom_bar(stat= “identity”)

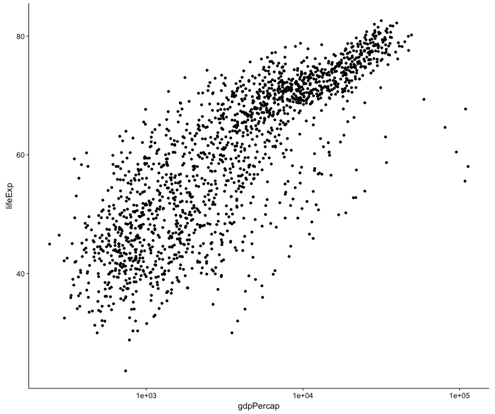
Grammar of Graphics

- Layered approach
- Iterative
- Customizable

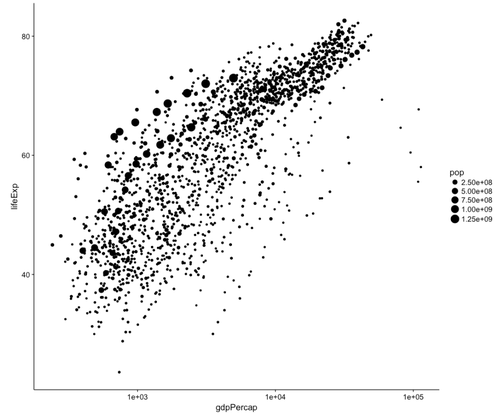
```
# A tibble: 1,704 x 6
       country continent  year lifeExp      pop gdpPercap
        <fctr>    <fctr> <int>   <dbl>    <int>     <dbl>
1  Afghanistan      Asia  1952  28.801  8425333  779.4453
2  Afghanistan      Asia  1957  30.332  9240934  820.8530
3  Afghanistan      Asia  1962  31.997 10267083  853.1007
4  Afghanistan      Asia  1967  34.020 11537966  836.1971
5  Afghanistan      Asia  1972  36.088 13079460  739.9811
```



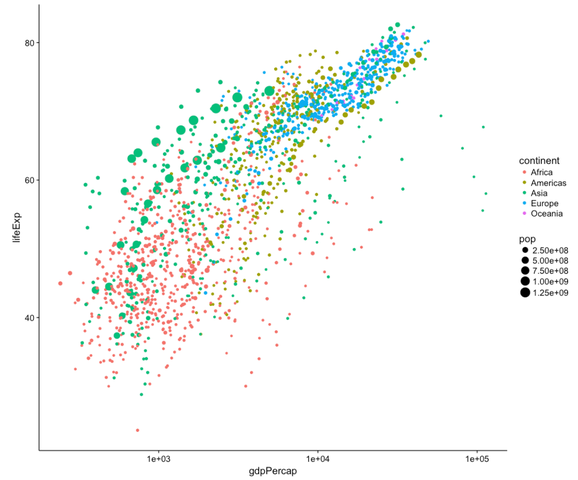

```
data %>%
  ggplot(aes(x = gdpPercap, y = lifeExp))+
  geom_point()
```




```
data %>%
  ggplot(aes(x = gdpPercap, y = lifeExp))+
  geom_point()+
  scale_x_log10()
```




```
data %>%
  ggplot(aes(x = gdpPercap, y = lifeExp, size = pop))+
  geom_point()+
  scale_x_log10()
```




```
data %>%
  ggplot(aes(x = gdpPercap, y = lifeExp,
             size = pop, color = continent))+
  geom_point()+
  scale_x_log10()
```

Demo!

Try to plot:

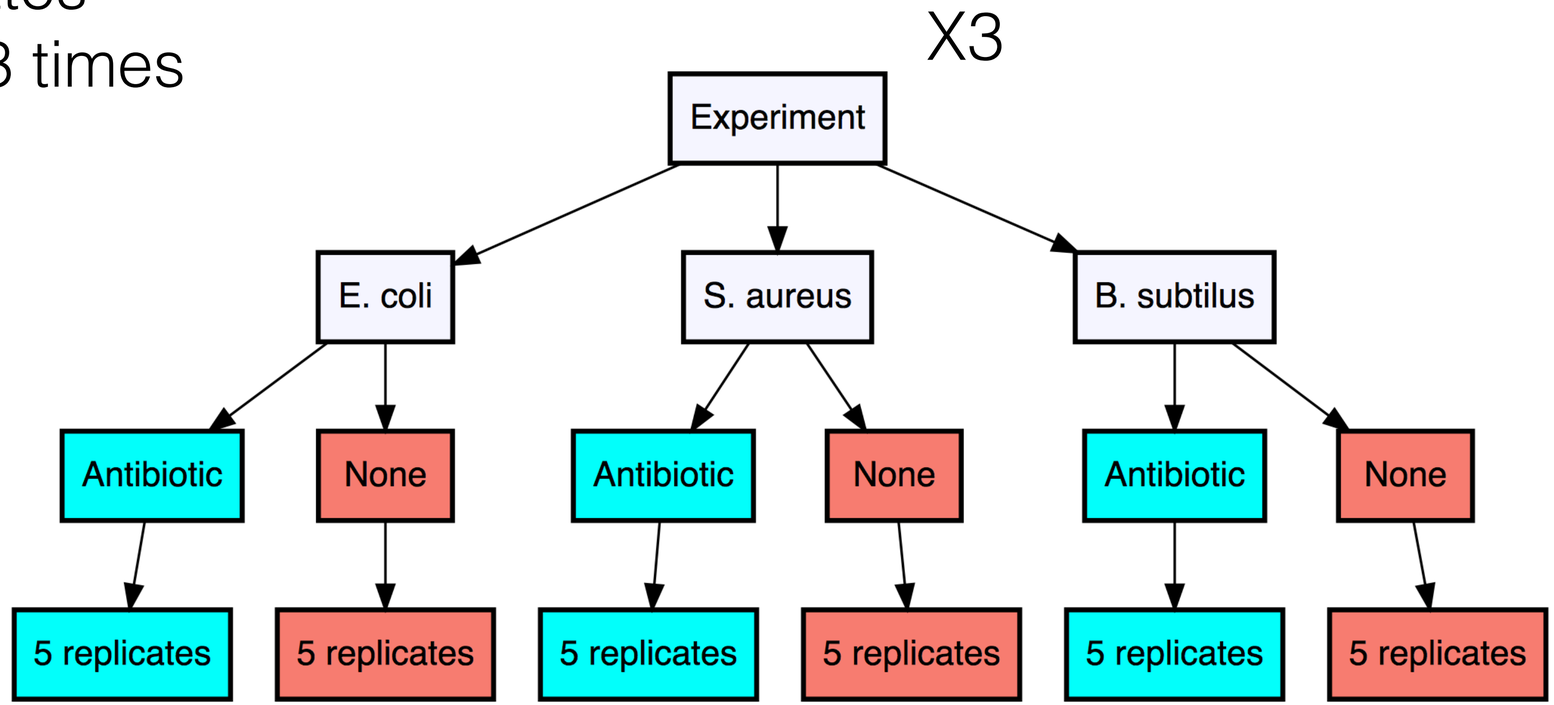
- Life expectancy of Asian countries in 1992
- Life expectancy of of Africa and Europe in 2007
- GDP of Americas and Europe in 2002

ggplot2

- Tweaking your plot
  - reorder - ggplot(aes (x = reorder(_____ , ordered_variable), y = ____)
  - color - + scale_fill_brewer(palette= “YlOrRd”)
  - legends - remove legend + guides(color = FALSE)/guides(fill = FALSE)
  - annotations - + annotate(“text”,x= 2, y=2, label = “your text”)

Experimental Design

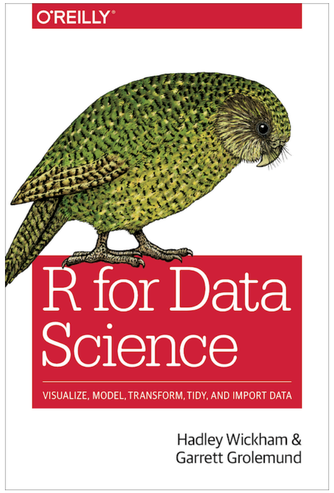
- **Three organisms** – E. coli, S. aureus, B. subtilus
- **Two treatments** - Antibiotic, None
- **Experiment** - 5 replicates
- **Repeat experiment** - 3 times



Additional Resources

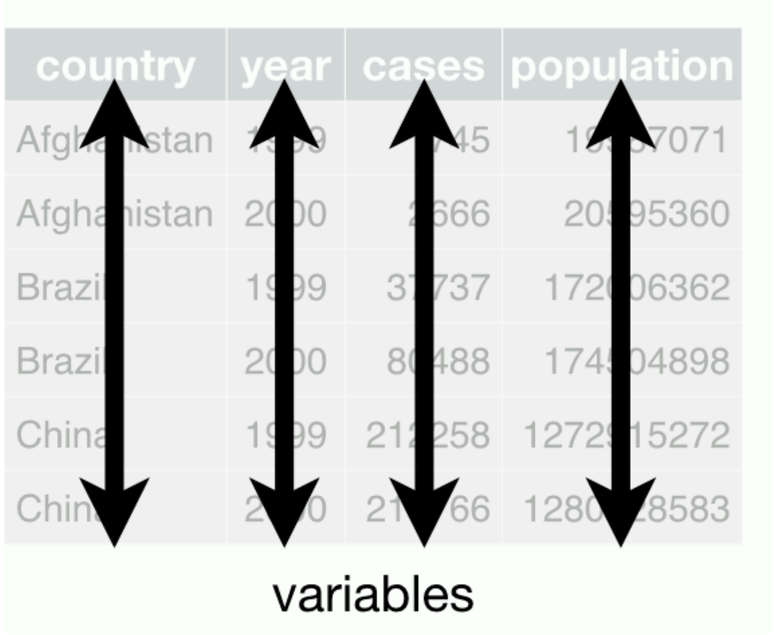
R Graphics Cookbook

R for Data Science

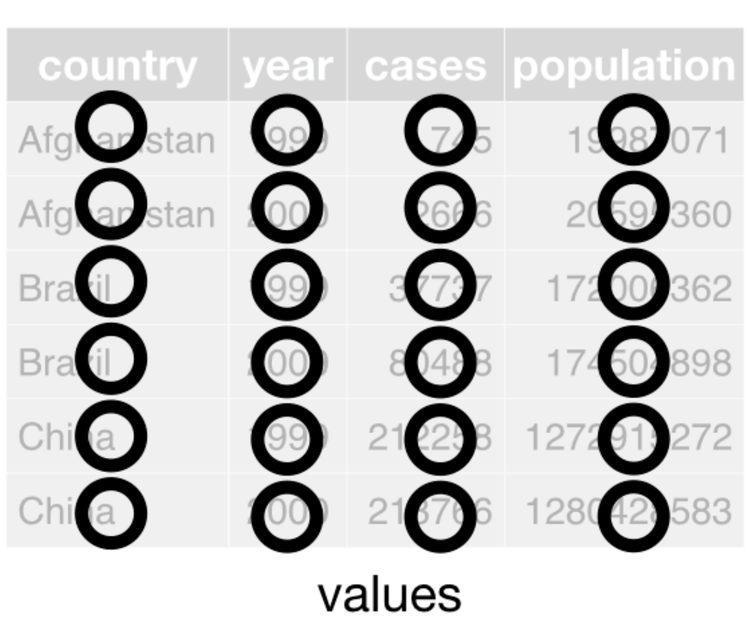


Stack Overflow

Tidy data


country
year
cases
population
variables


country
year
cases
population
observations


country
year
cases
population
values

Wide format

Treatment	1_Ecoli	1_Saureus	1_Bsubtilis	2_Ecoli	2_Saureus	2_Bsubtilis
Antibiotic	285	240	312	362	244	415
Antibiotic	345	371	461	368	375	315
Antibiotic	298	337	352	287	228	370
Antibiotic	286	394	494	378	302	314
Antibiotic	354	213	311	363	349	303
None	146	286	340	228	284	363
None	180	300	285	246	262	381
None	137	279	271	166	266	325
None	179	253	355	226	270	398
None	168	272	424	175	258	336

Long format (tidy)

Treatment	Experiment	Organism	Count
Antibiotic	1	Ecoli	285
Antibiotic	1	Ecoli	345
Antibiotic	1	Ecoli	298
Antibiotic	1	Ecoli	286
Antibiotic	1	Ecoli	354
None	1	Ecoli	146
None	1	Ecoli	180
None	1	Ecoli	137
None	1	Ecoli	179
None	1	Ecoli	168

Tidy data

Treatment	Experiment	Organism	Count
Antibiotic	1	Ecoli	285
Antibiotic	1	Ecoli	345
Antibiotic	1	Ecoli	298
Antibiotic	1	Ecoli	286
Antibiotic	1	Ecoli	354
None	1	Ecoli	146
None	1	Ecoli	180
None	1	Ecoli	137
None	1	Ecoli	179
None	1	Ecoli	168

Organism	Treatment	Experiment	N	mean	sd	se
Ecoli	Antibiotic	1	5	313.6	33.32116445	14.90167776
Ecoli	Antibiotic	2	5	351.6	36.66469692	16.39695094
Ecoli	Antibiotic	3	5	346.2	44.80736547	20.03846301
Ecoli	None	1	5	162	19.55760722	8.746427842
Ecoli	None	2	5	208.2	35.42880184	15.84424186
Ecoli	None	3	5	177.6	40.14722905	17.95438665