190F Foundations of Data Science

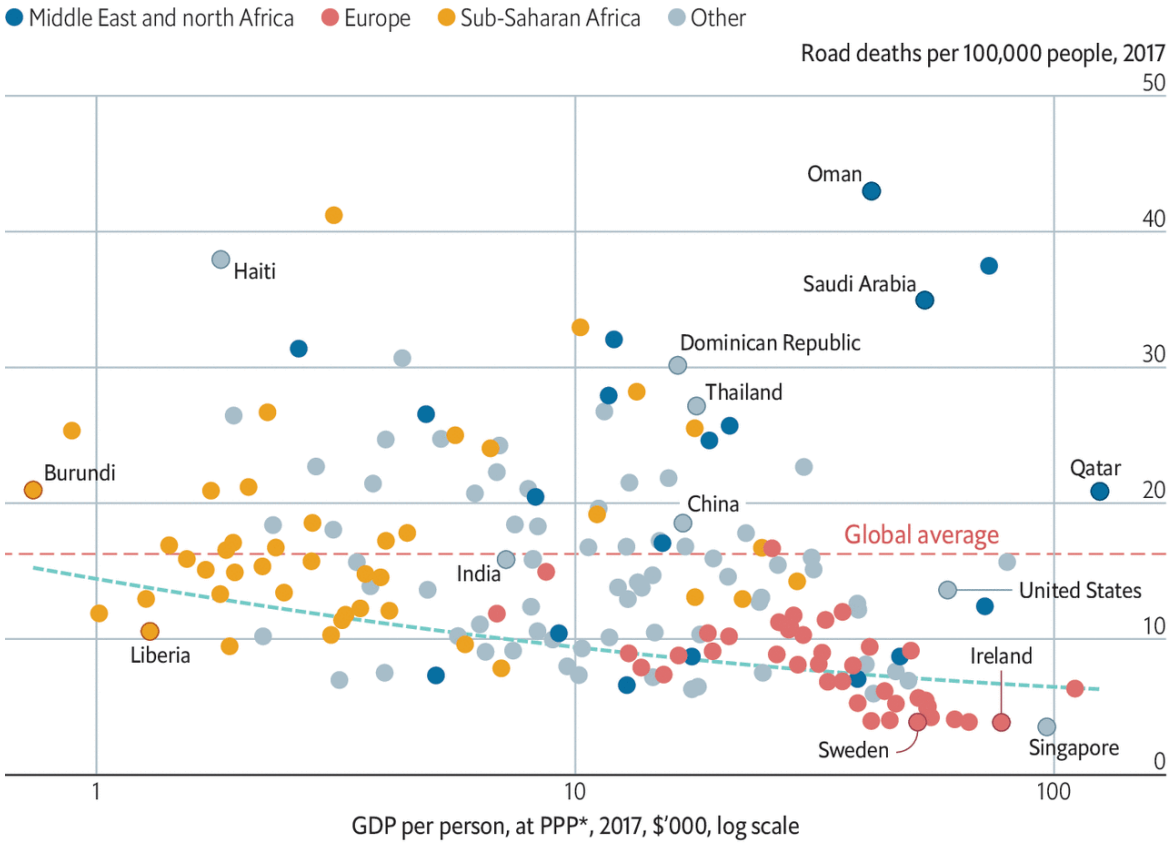
Spring 2020

Lecture 3

Expressions
& Building Tables

Announcements

Driving lesson



Sources: World Bank; IHME

*Purchasing-power parity

The Economist

Arithmetic

Arithmetic Operators

Operation	Operator	Example	Value
Addition	+	2 + 3	5
Subtraction	-	2 - 3	-1
Multiplication	*	2 * 3	6
Division	/	7 / 3	2.66667
Remainder	%	7 % 3	1
Exponentiation	**	2 ** 0.5	1.41421

(Demo)

Ints and Floats

Python has two real number types

- **`int`**: an integer of any size
- **`float`**: a number with an optional fractional part

An **`int`** never has a decimal point; a **float** always does

A **`float`** might be printed using scientific notation

Three limitations of float values:

- They have limited size (but the limit is huge)
- They have limited precision of 15-16 decimal places
- After arithmetic, the final few decimal places can be wrong

Arithmetic Question

Rank the results of the following expressions in order from least to greatest

```
A.3 * 10 ** 10
B.10 * 3 ** 10
C.(10 * 3) ** 10
D.10 / 3 / 10
E.10 / (3 / 10)
```

```
A.30000000000
B.590490
C.590490000000000
D.0.33333333333333337
E.33.333333333333336
```

Strings

Text and Strings

A string value is a snippet of text of any length

- `'a'`
- `'word'`
- `"there can be 2 sentences. Here's the second!"`

Strings that contain numbers can be converted to numbers

- `int('12')`
- `float('1.2')`

Any value can be converted to a string

- `str(5)`

(Demo)

Discussion Question

Assume you have run the following statements

```
x = 3
y = '4'
z = '5.6'
```

What's the source of the error in each example?

A.
```
x + y
```

B.
```
x + int(y + z)
```

C.
```
str(x) + int(y)
```

D.
```
str(x, y) + z
```

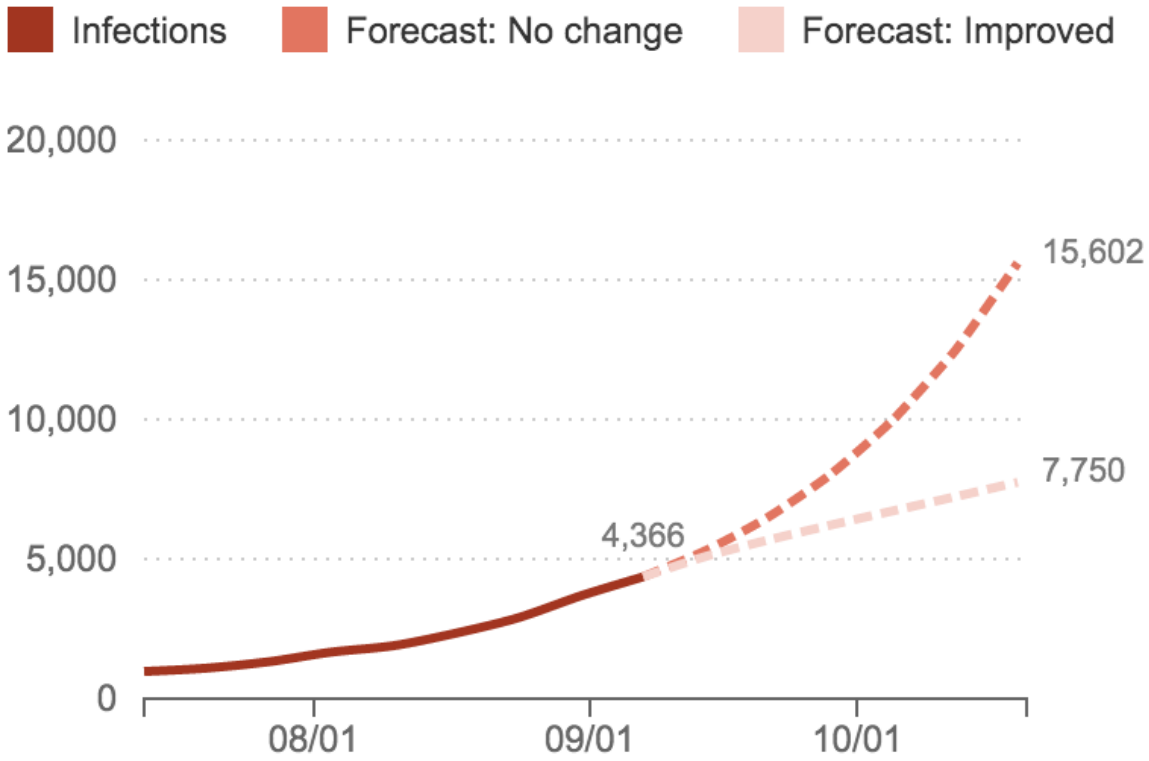
Exponential Growth

Ebola Epidemic, Sept. 2014

A Frightening Curve: How Fast Is The Ebola Outbreak Growing?

"It's spreading and growing *exponentially*," President Obama said.

"This is a disease outbreak that is advancing in an exponential fashion," said Dr. David Nabarro, who is heading the U.N.'s effort against Ebola.



http://www.npr.org/sections/goatsandsoda/2014/09/18/349341606/why-the-math-of-the-ebola-epidemic-is-so-scary

Source: Columbia Prediction of Infectious Diseases, World Health Organization

Growth Rate

- The rate of increase per unit time
- After one time unit, a quantity **`x`** growing at rate **`g`** will be

```
x * (1 + g)
```

- After **`t`** time units, a quantity **`x`** growing at rate **`g`** will be

```
x * (1 + g) ** t
```

- If **`after`** and **`before`** are measurements of the same quantity taken **`t`** time units apart, then the *growth rate* is

```
(after/before) ** (1/t) - 1
```

(Demo)

Arrays

Arrays

An array contains a sequence of values

- All elements of an array should have the same type
- Arithmetic is applied to each element individually
- When two arrays are added, they must have the same size; corresponding elements are added in the result
- A column of a table is an array

(Demo)

Ranges

Ranges

A range is an array of consecutive numbers

- **`np.arange(end)`**:
  An array of increasing integers from 0 up to **`end`**
- **`np.arange(start, end)`**:
  An array of increasing integers from **`start`** up to **`end`**
- **`np.arange(start, end, step)`**:
  A range with **`step`** between consecutive values

The range always includes **`start`** but excludes **`end`**

Tables Review

Ways to create a table

- `Table.read_table(filename)` - reads a table from a spreadsheet
- `Table()` - an empty table
- and...

Arrays → Tables

- `Table().with_column(label, data)` - creates a table with a single column; `data` is an array
- `Table().with_columns(label1, data1, ...)` - creates a table, with an array of data for each column

Table Methods

- Creating and extending tables:
  - **`Table().with_columns`** and **`Table.read_table`**
- Finding the size: **`num_rows`** and **`num_columns`**
- Referring to columns: labels, relabeling, and indices
  - **`labels`** and **`relabeled`**; column indices start at 0
- Accessing data in a column
  - **`column`** takes a label or index and returns an array
- Using array methods to work with data in columns
  - **`item`**, **`sum`**, **`min`**, **`max`**, and so on
- Creating new tables containing some of the original columns:
  - **`select`**, **`drop`**

Examples

The table **`students`** has columns **`Name`**, **`ID`**, and **`Score`**.

Write one line of code that evaluates to:

a) A table consisting of only the column labeled **`Name`**

```
students.select('Name')
```

b) The largest score

```
students.column('Score').max()
max(students.column('Score'))
```