Week 7: Comparisons, Compositions, and Time

Spring 2017

Matthew Turk

Broadcasting

go.ischool.illinois.edu/meet2

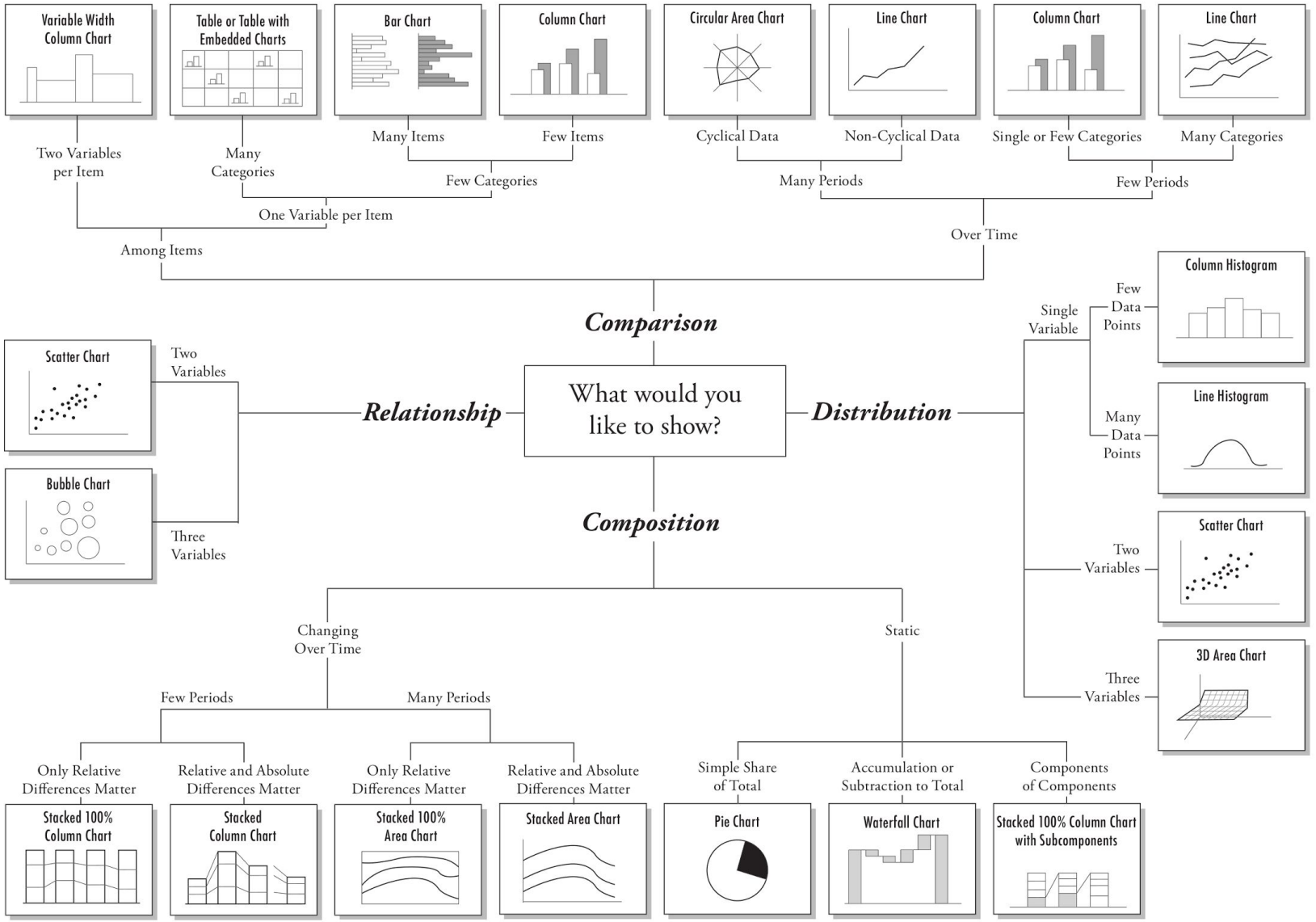
Last Week

- How Colors Work
  - HSV / RGB / Eyeballs
  - Rotations, perception
- Palette types
  - Sequential
  - Qualitative
  - Diverging

Today

- Types of Visualizations: Part II
  - “If this, then that…”
- Composition plots
  - Examples
  - How-to
- Comparison plots
  - Examples
  - How-to
- Pandas

Chart Suggestions—A Thought-Starter

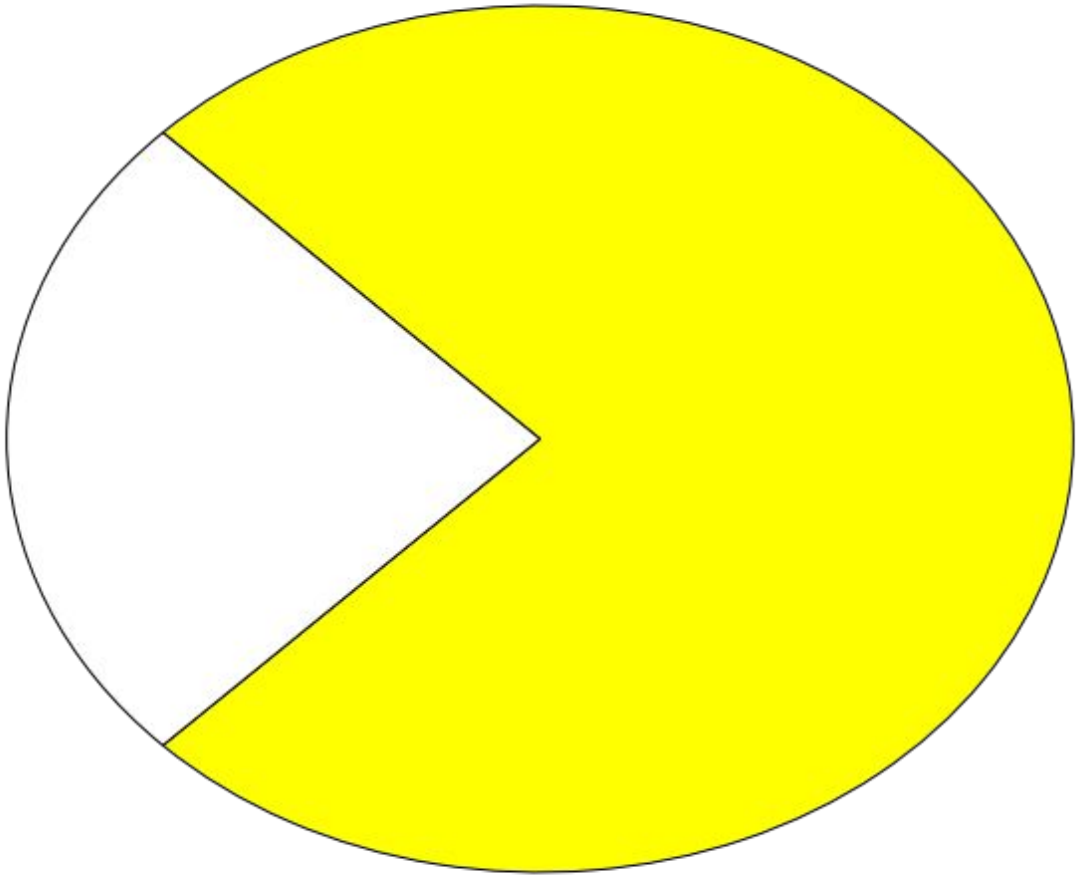


Credit: Andrew Abela

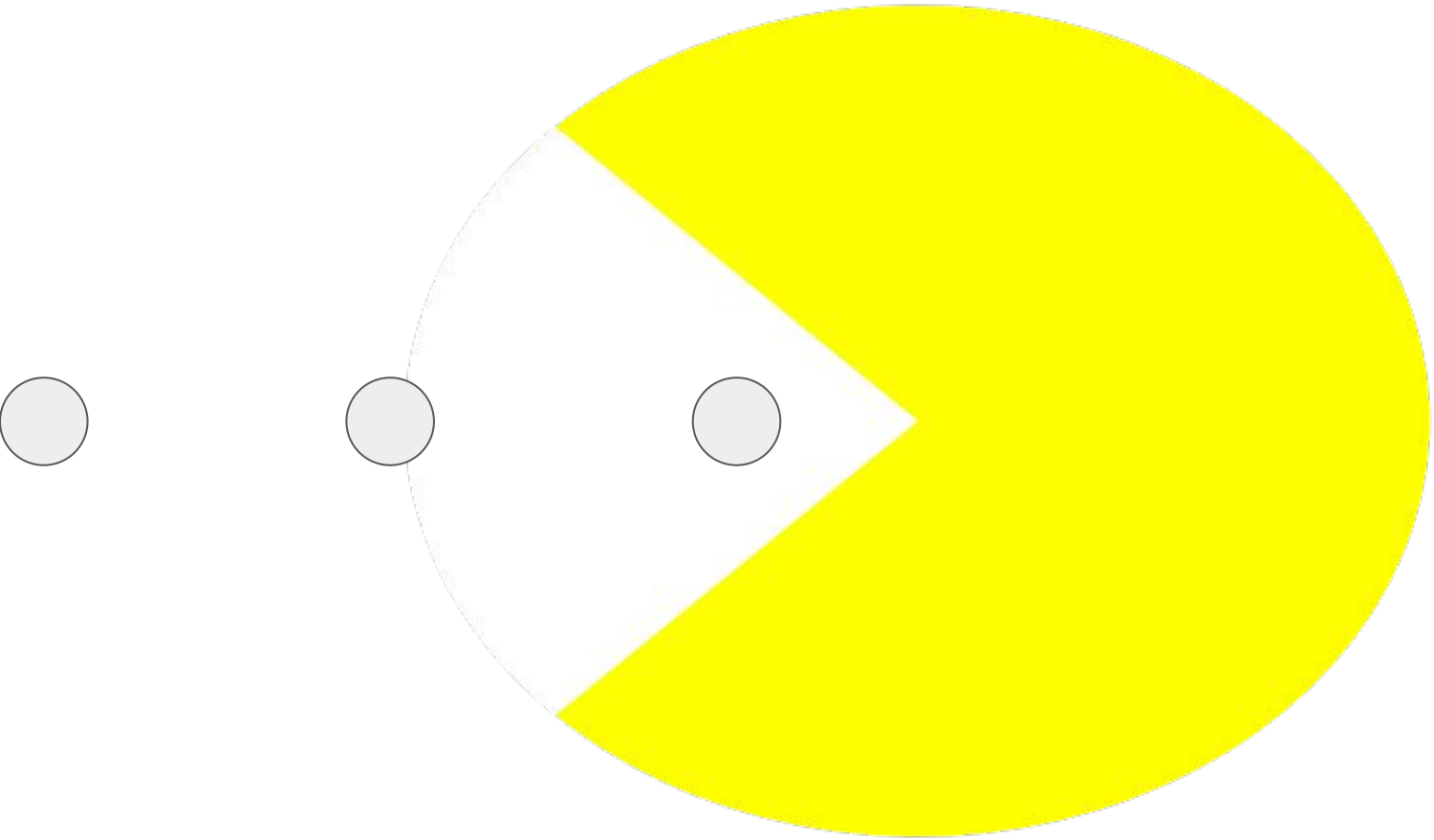
Composition Plots

- Static
  - Pie Chart
  - Area chart
- Changing over time

Pie Chart

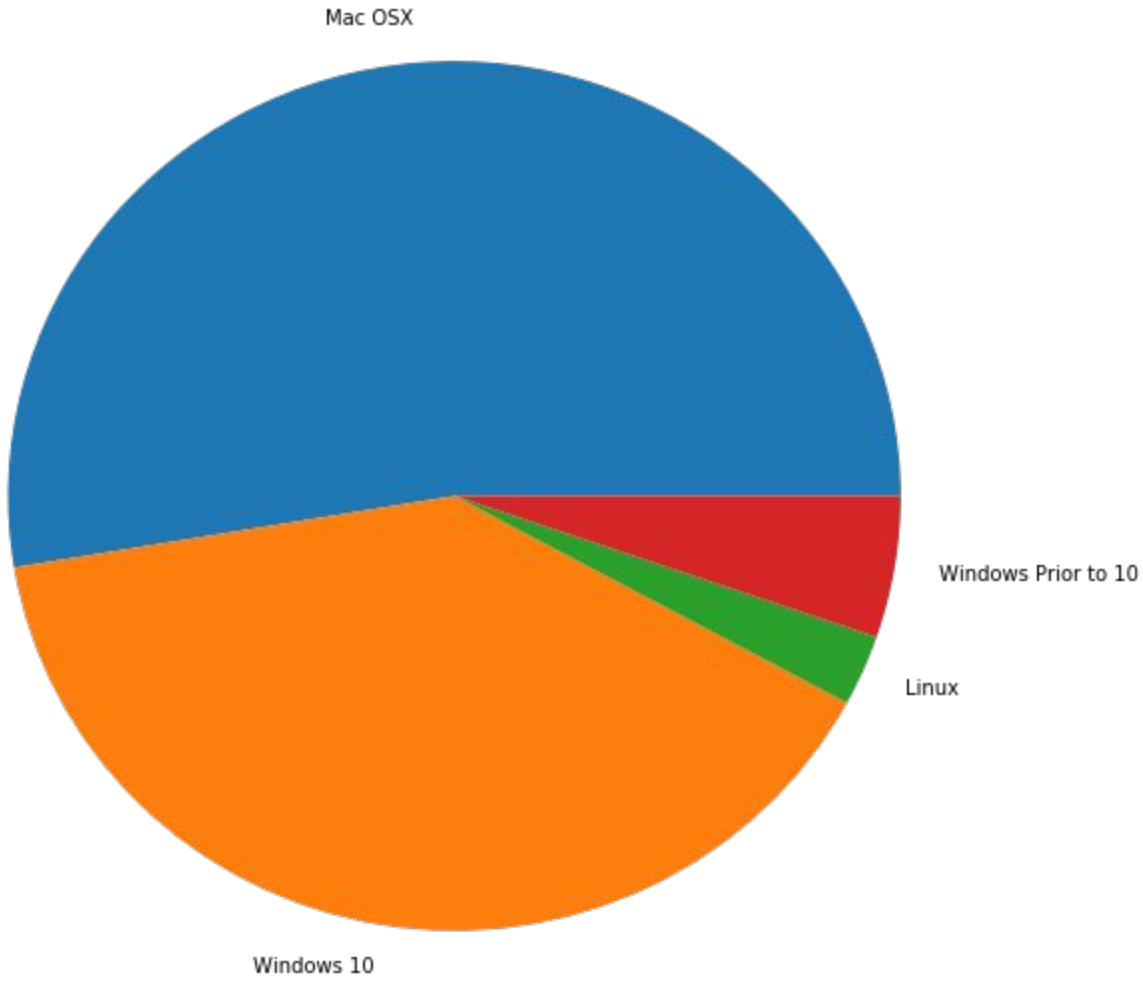


Pie Chart

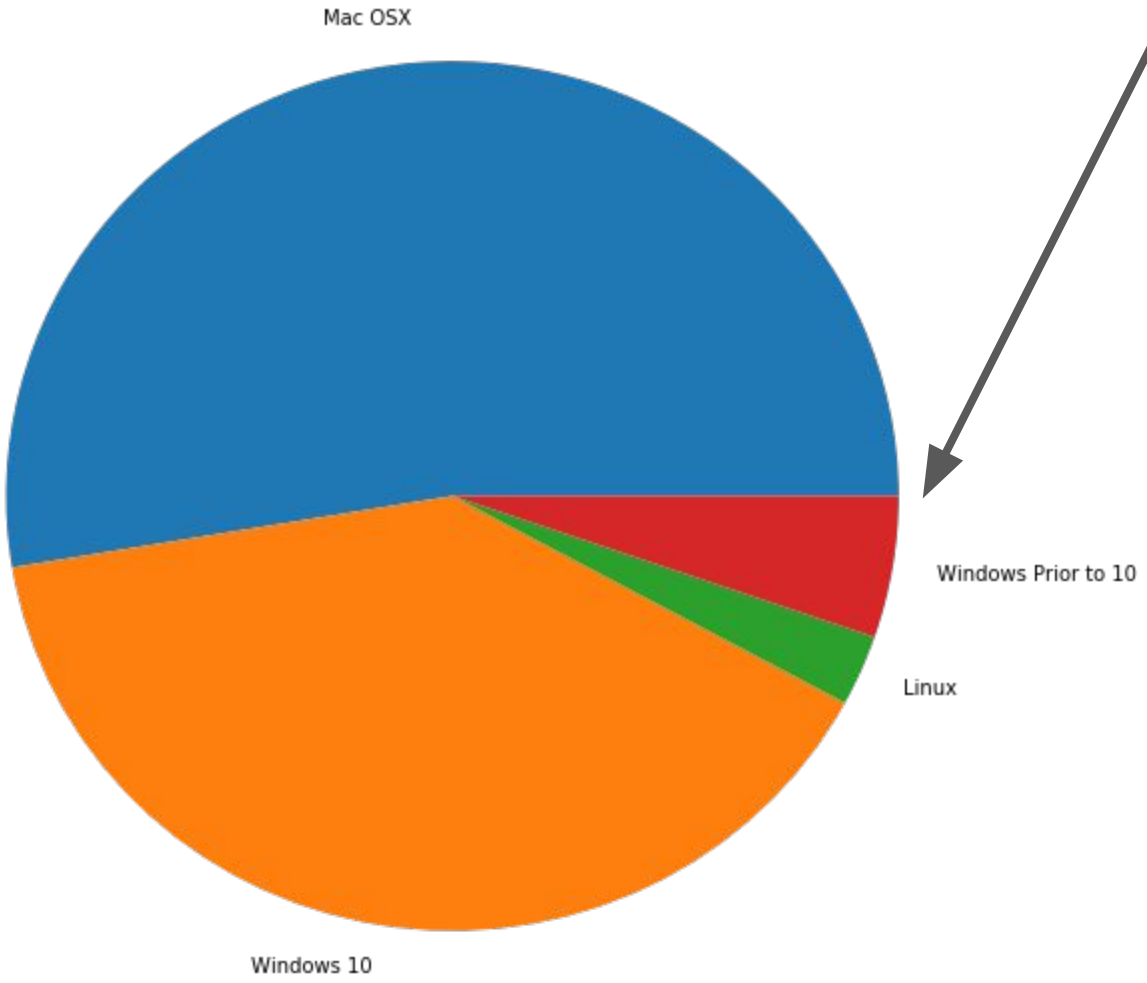


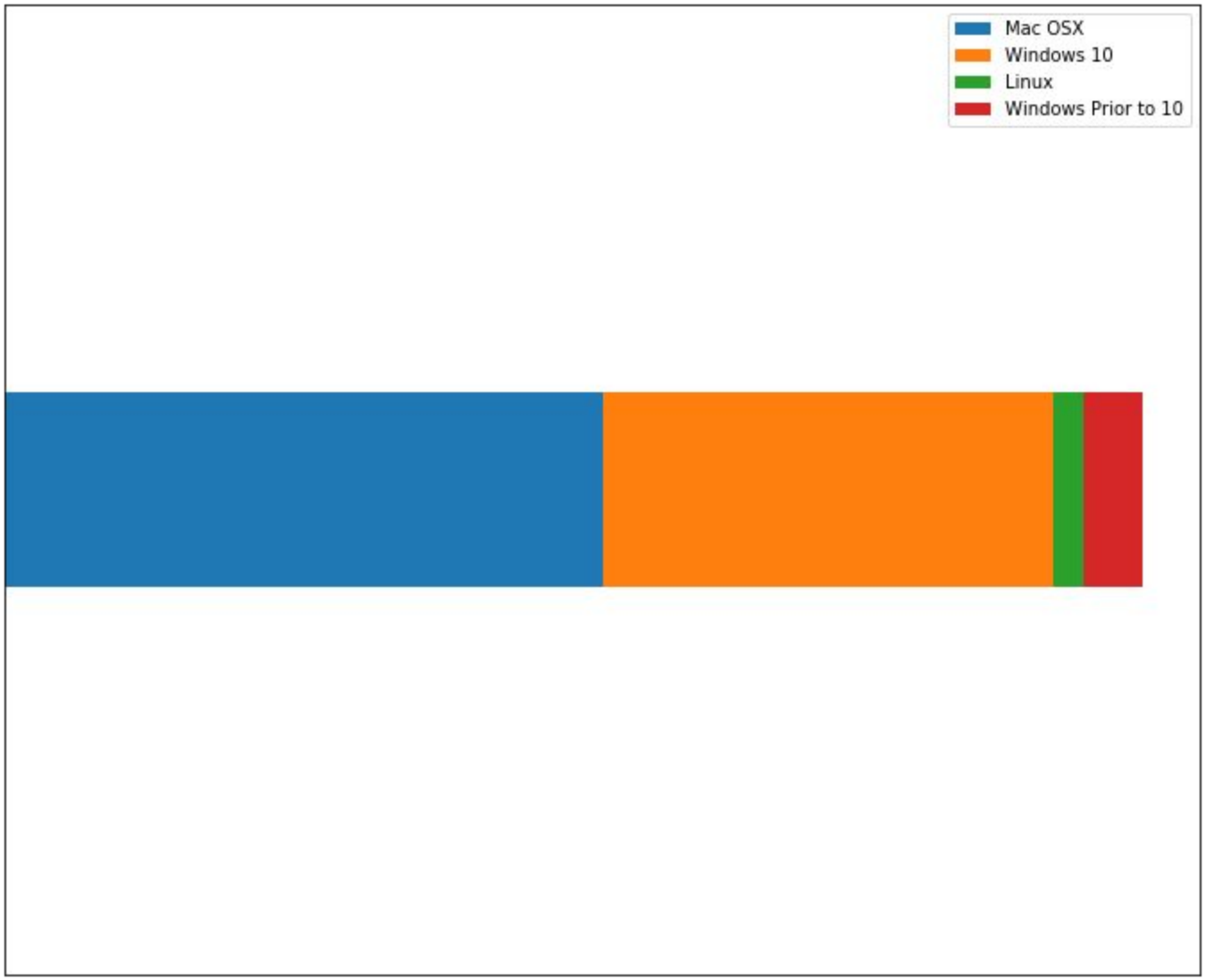
You probably don't
want a pie chart.

http://www.businessinsider.com/pie-charts-are-the-worst-2013-6

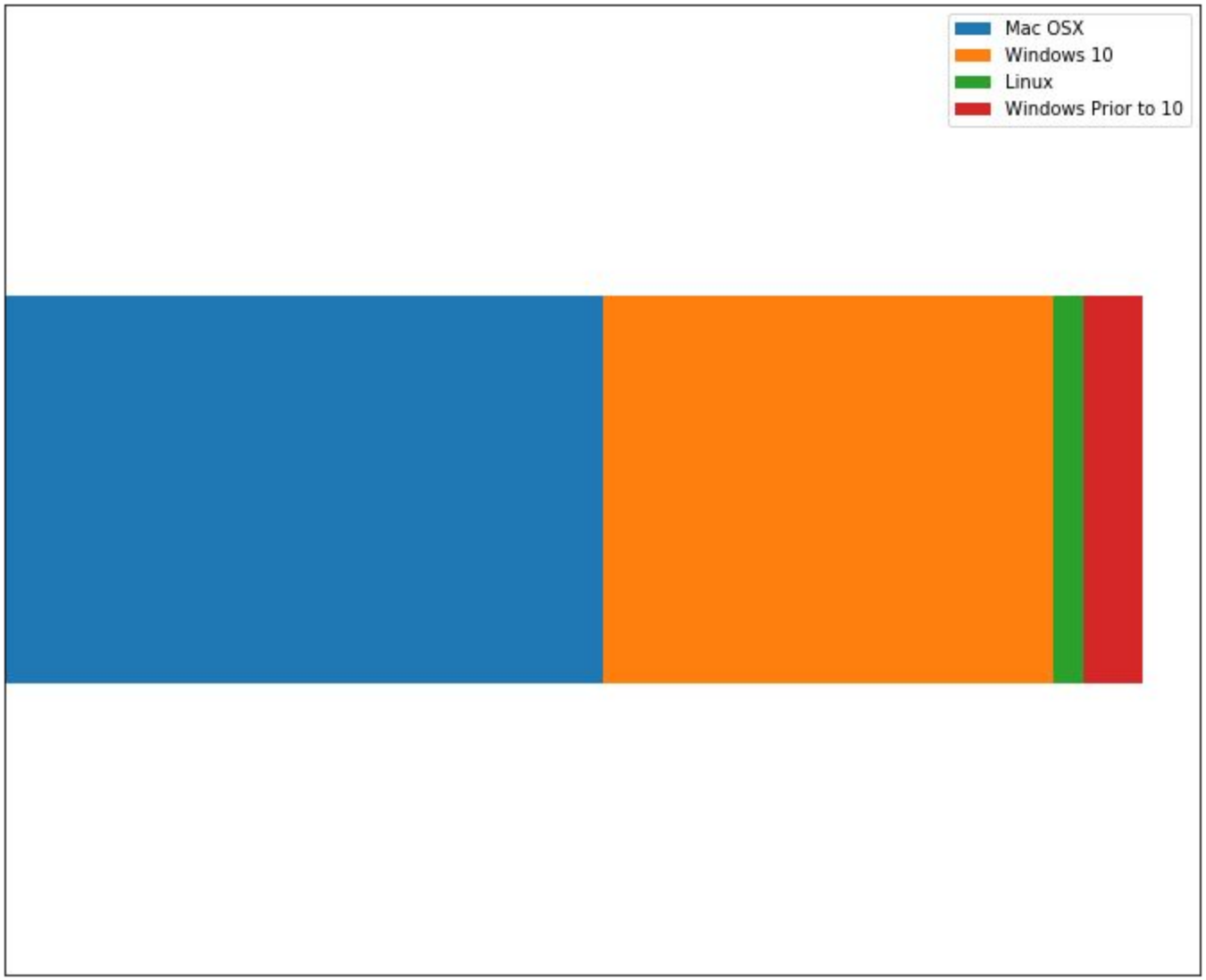

Mac OSX
Windows Prior to 10
Linux
Windows 10

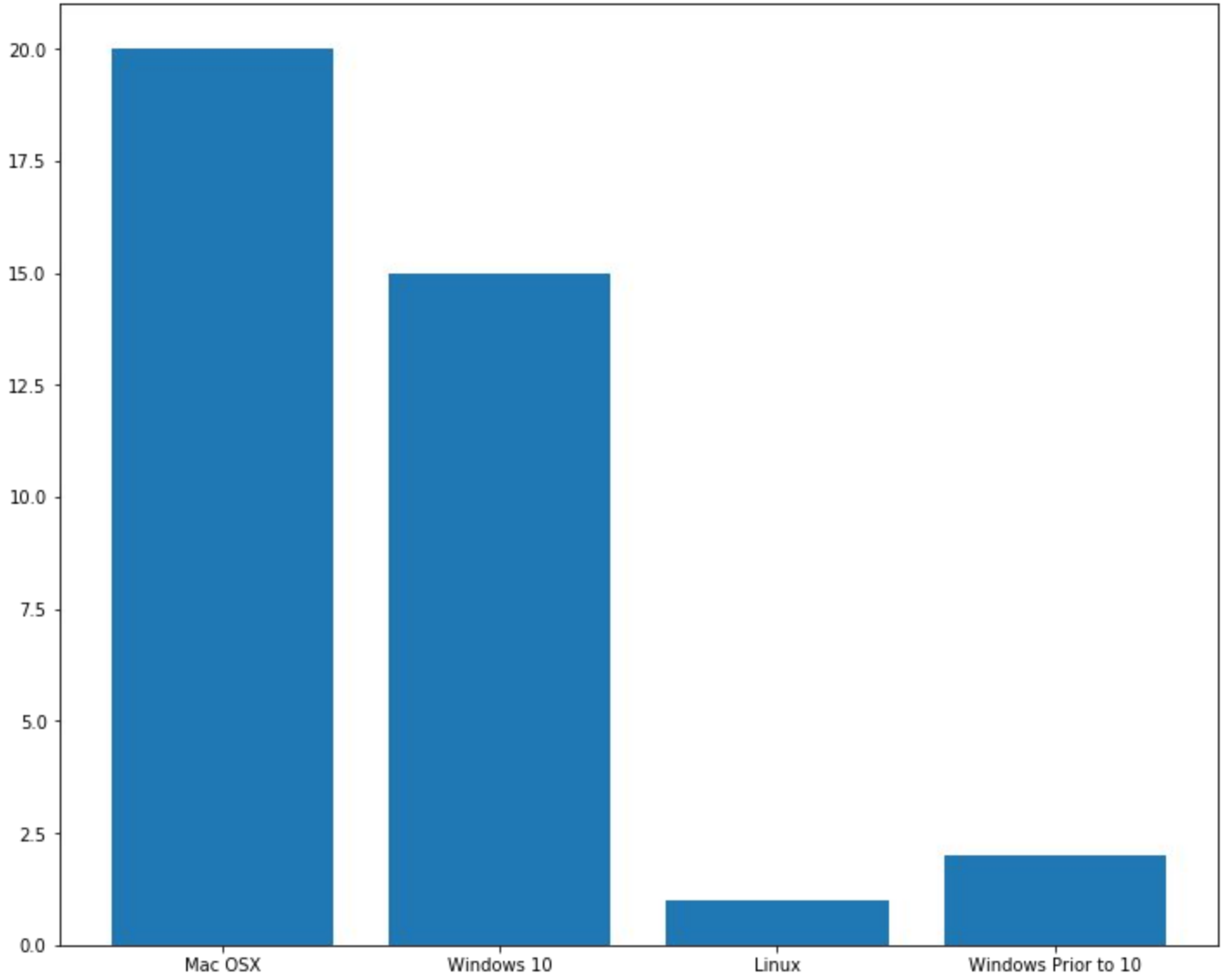
Essentially impossible to gauge relative size of non-majority wedges



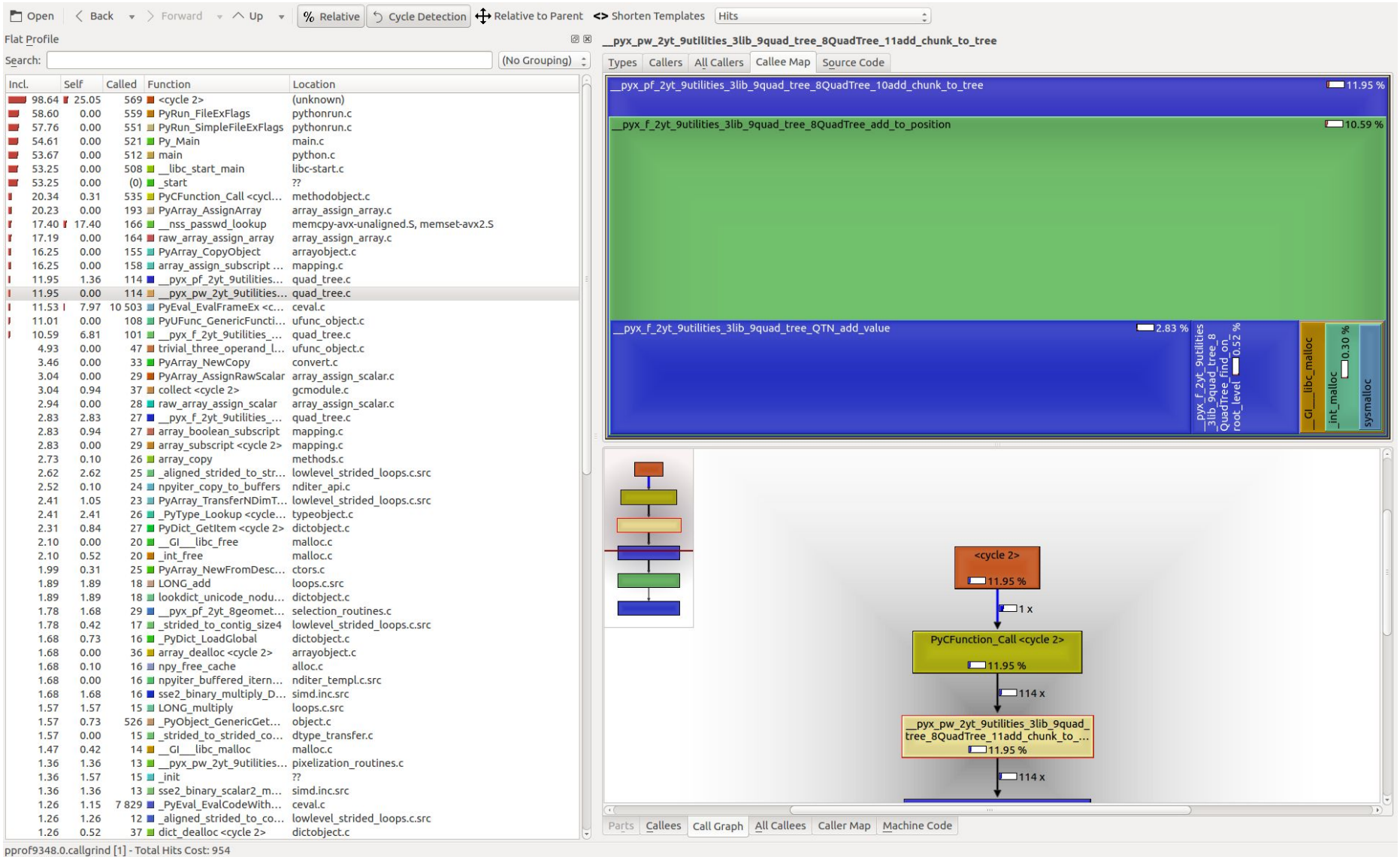

Mac OSX
Windows 10
Linux
Windows Prior to 10

Same data, different vertical zoom.




20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Mac OSX
Windows 10
Linux
Windows Prior to 10

Area Chart


Open
Back
Forward
Up
% Relative
Cycle Detection
Relative to Parent
<> Shorten Templates
Hits
Flat Profile
Search:
(No Grouping)
Incl. Self Called Function Location
98.64 25.05 569 <cycle 2> (unknown)
58.60 0.00 559 PyRun_FileExFlags pythonrun.c
57.76 0.00 551 PyRun_SimpleFileExFlags pythonrun.c
54.61 0.00 521 Py_Main main.c
53.67 0.00 512 main python.c
53.25 0.00 508 __libc_start_main libc-start.c
53.25 0.00 (0) _start ??
20.34 0.31 535 PyCFunction_Call <cycl... methodobject.c
20.23 0.00 193 PyArray_AssignArray array_assign_array.c
17.40 17.40 166 __nss_passwd_lookup memcpy-avx-unaligned.S, memset-avx2.S
17.19 0.00 164 raw_array_assign_array array_assign_array.c
16.25 0.00 155 PyArray_CopyObject arrayobject.c
16.25 0.00 158 array_assign_subscript ... mapping.c
11.95 1.36 114 __pyx_pf_2yt_9utilities... quad_tree.c
11.95 0.00 114 __pyx_pw_2yt_9utilities... quad_tree.c
11.53 7.97 10 503 PyEval_EvalFrameEx <c... ceval.c
11.01 0.00 108 PyUFunc_GenericFuncti... ufunc_object.c
10.59 6.81 101 __pyx_f_2yt_9utilities_... quad_tree.c
4.93 0.00 47 trivial_three_operand_l... ufunc_object.c
3.46 0.00 33 PyArray_NewCopy convert.c
3.04 0.00 29 PyArray_AssignRawScalar array_assign_scalar.c
3.04 0.94 37 collect <cycle 2> gcmodule.c
2.94 0.00 28 raw_array_assign_scalar array_assign_scalar.c
2.83 2.83 27 __pyx_f_2yt_9utilities_... quad_tree.c
2.83 0.94 27 array_boolean_subscript mapping.c
2.83 0.00 29 array_subscript <cycle 2> mapping.c
2.73 0.10 26 array_copy methods.c
2.62 2.62 25 _aligned_strided_to_str... lowlevel_strided_loops.c.src
2.52 0.10 24 npyiter_copy_to_buffers nditer_api.c
2.41 1.05 23 PyArray_TransferNDimT... lowlevel_strided_loops.c.src
2.41 2.41 26 _PyType_Lookup <cycle... typeobject.c
2.31 0.84 27 PyDict_GetItem <cycle 2> dictobject.c
2.10 0.00 20 __GI___libc_free malloc.c
2.10 0.52 20 _int_free malloc.c
1.99 0.31 25 PyArray_NewFromDesc... ctors.c
1.89 1.89 18 LONG_add loops.c.src
1.89 1.89 18 lookdict_unicode_nodu... dictobject.c
1.78 1.68 29 __pyx_pf_2yt_8geomet... selection_routines.c
1.78 0.42 17 _strided_to_contig_size4 lowlevel_strided_loops.c.src
1.68 0.73 16 _PyDict_LoadGlobal dictobject.c
1.68 0.00 36 array_dealloc <cycle 2> arrayobject.c
1.68 0.10 16 npy_free_cache alloc.c
1.68 0.00 16 npyiter_buffered_itern... nditer_templ.c.src
1.68 1.68 16 sse2_binary_multiply_D... simd.inc.src
1.57 1.57 15 LONG_multiply loops.c.src
1.57 0.73 526 _PyObject_GenericGet... object.c
1.57 0.00 15 _strided_to_strided_co... dtype_transfer.c
1.47 0.42 14 __GI___libc_malloc malloc.c
1.36 1.36 13 __pyx_pw_2yt_9utilities... pixelization_routines.c
1.36 1.57 15 _init ??
1.36 1.36 13 sse2_binary_scalar2_m... simd.inc.src
1.26 1.15 7 829 _PyEval_EvalCodeWith... ceval.c
1.26 1.26 12 _aligned_strided_to_co... lowlevel_strided_loops.c.src
1.26 0.52 37 dict_dealloc <cycle 2> dictobject.c
__pyx_pw_2yt_9utilities_3lib_9quad_tree_8QuadTree_11add_chunk_to_tree
Types
Callers
All Callers
Callee Map
Source Code
__pyx_pf_2yt_9utilities_3lib_9quad_tree_8QuadTree_10add_chunk_to_tree
11.95 %
__pyx_f_2yt_9utilities_3lib_9quad_tree_8QuadTree_add_to_position
10.59 %
__pyx_f_2yt_9utilities_3lib_9quad_tree_QTN_add_value
2.83 %
__pyx_f_2yt_9utilities_3lib_9quad_tree_8QuadTree_find_on_root_level
0.52 %
__GI___libc_malloc
_int_malloc
0.30 %
sysmalloc
<cycle 2>
11.95 %
1 x
PyCFunction_Call <cycle 2>
11.95 %
114 x
__pyx_pw_2yt_9utilities_3lib_9quad_tree_8QuadTree_11add_chunk_to_...
11.95 %
114 x
Parts
Callees
Call Graph
All Callees
Caller Map
Machine Code
pprof9348.0.callgrind [1] - Total Hits Cost: 954

Comparison Plots

- Among Items
  - One Variable, Few Categories: Column, or collection of bars
  - Two Variables: Variable Width Column Chart
  - Many variables: Embedded table or charts
- Changing Over Time
  - Many Periods, non-cyclical: Line chart
  - Few Periods: Column or Line (depending on number of categories)

Pandas

- pandas.pydata.org
- Support for indexing, multi-indexing
- Data structures
  - Series
  - DataFrame
  - Panel
- Groupby, select, aggregation features
- IO features
  - Reading/writing CSV, HDF5
  - Loading data from remote sources

Data Types

- Categorical data
- Date time
- Indexing by columns or rows

What We Will Learn

- How to read in data
- How to index and select that data
- How to replace values and columns
- How to plot

Core Concepts

- Columns
- Group-by
- Aggregate
- Index / locate
- dtypes and categorical types

Case Study: World Bank Indicators

```
{'NE.GDI.FPRV.CN': 'Gross fixed capital formation, private sector (current LCU)',
 'NE.CON.PETC.CN': 'Household final consumption expenditure, etc. (current LCU)',
 'SP.POP.TOTL': 'Population, total',
 'EN.ATM.NOXE.EG.ZS': 'Nitrous oxide emissions in energy sector (% of total)',
 'NE.GDI.FTOT.KD': 'Gross fixed capital formation (constant 2010 US$)',
 'NY.GNP.MKTP.CN': 'GNI (current LCU)',
 'SH.STA.TRAF.P5': 'Mortality caused by road traffic injury (per 100,000 people)',
  ...}
```