lec13

September 29, 2021

[1]:
```
from datascience import *
import numpy as np

%matplotlib inline
import matplotlib.pyplot as plots
plots.style.use('fivethirtyeight')
```

0.1 Comparison

[2]:
```
3 > 1
```

[2]: True

[3]:
```
type(3 > 1)
```

[3]: bool

[4]:
```
3 < 1
```

[4]: False

[5]:
```
True
```

[5]: True

[6]:
```
3 == 3
```

[6]: True

[7]:
```
3 = 3
```

```
  File "<ipython-input-7-79bfd1be65e2>", line 1
    3 = 3
    ^
SyntaxError: cannot assign to literal
```

[8]:
```
x = 14
y = 3
```

[9]:
```
x > 10 and y > 5
```

[9]:
```
False
```

0.2 Comparisons with arrays

[10]:
```
pets = make_array('cat', 'dog', 'cat', 'cat', 'dog', 'rabbit')
```

[11]:
```
pets == 'dog'
```

[11]:
```
array([False,  True, False, False,  True, False])
```

[12]:
```
pets > 'cat'
```

[12]:
```
array([False,  True, False, False,  True,  True])
```

[13]:
```
sum(make_array(False, True, False, False, True, False))
```

[13]:
```
2
```

[14]:
```
sum(pets == 'dog')
```

[14]:
```
2
```

[15]:
```
np.count_nonzero(pets == 'dog')
```

[15]:
```
2
```

0.3 Advanced `where` with predicates to pick rows

[16]:
```
terms = Table().with_column('Semester', np.arange(1, 9))
terms
```

[16]:
```
Semester
1
2
3
4
5
6
7
8
```

[17]:
```
terms.where('Semester', are.above(6))
```

[17]:
```
Semester
7
8
```

[18]:
```
def is_senior(x):
    return x > 6
```

[19]:
```
is_senior(5)
```

[19]:
```
False
```

[20]:
```
terms.apply(is_senior, 'Semester')
```

[20]:
```
array([False, False, False, False, False, False,  True,  True])
```

[21]:
```
terms.where('Semester', are.above(6))
```

[21]:
```
Semester
7
8
```

[22]:
```
terms.where(terms.apply(is_senior, 'Semester'))
```

[22]:
```
Semester
7
8
```

0.4 Conditional Statements

[23]:
```
def year_from_semesters(x):
    if x <= 0:
        return 'Not a valid input'
    elif x <= 2:
        return 'Freshman'
    elif x <= 4:
        return 'Sophomore'
    elif x <= 6:
        return 'Junior'
    elif x <= 8:
        return 'Senior'
```

[24]:
```
year_from_semesters(-15.6)
```

[24]:
```
'Not a valid input'
```

[25]:
```
year_from_semesters(5)
```

[25]:
```
'Junior'
```

[26]:
```
year_from_semesters(9001)
```

[27]:
```
terms.with_column('Year', terms.apply(year_from_semesters, 'Semester'))
```

[27]:
```
Semester | Year
1        | Freshman
2        | Freshman
3        | Sophomore
4        | Sophomore
5        | Junior
6        | Junior
7        | Senior
8        | Senior
```

0.5 Appending Arrays

[28]:
```
first = np.arange(1, 6)
second = np.arange(6, 11)
```

[29]:
```
first
```

[29]:
```
array([1, 2, 3, 4, 5])
```

[30]:
```
second
```

[30]:
```
array([ 6,  7,  8,  9, 10])
```

[31]:
```
np.append(first, 30)
```

[31]:
```
array([ 1,  2,  3,  4,  5, 30])
```

[32]:
```
np.append(first, second)
```

[32]:
```
array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10])
```

[33]:
```
first
```

[33]:
```
array([1, 2, 3, 4, 5])
```

0.6 Random Selection

[34]:
```
two_groups = make_array('treatment', 'control')
```

[35]:
```
np.random.choice(two_groups)
```

[35]:
```
'control'
```

[36]:
```
np.random.choice(two_groups)
```

[36]:
```
'treatment'
```

[37]:
```
np.random.choice(two_groups, 10)
```

[37]:
```
array(['control', 'control', 'control', 'treatment', 'treatment',
       'control', 'treatment', 'treatment', 'treatment', 'treatment'],
      dtype='<U9')
```

[38]:
```
sum(np.random.choice(two_groups, 10) == 'treatment')
```

[38]:
```
7
```

[39]:
```
sum(np.random.choice(two_groups, 10) == 'control')
```

[39]:
```
6
```

[40]:
```
outcomes = np.random.choice(two_groups, 10)
```

[41]:
```
outcomes
```

[41]:
```
array(['control', 'control', 'control', 'control', 'treatment', 'control',
       'control', 'control', 'control', 'control'], dtype='<U9')
```

[42]:
```
sum(outcomes == 'treatment')
```

[42]:
```
1
```

[43]:
```
sum(outcomes == 'control')
```

[43]:
```
9
```

0.7 A simple experiment/simulation

[44]:
```
coin = ['heads', 'tails']
```

[45]:
```
np.random.choice(coin)
```

[45]:
```
'tails'
```

[46]:
```
tosses = np.random.choice(coin, 100)
tosses
```

[46]:
```
array(['tails', 'tails', 'heads', 'tails', 'tails', 'tails', 'heads',
       'heads', 'tails', 'heads', 'tails', 'tails', 'tails', 'tails',
       'tails', 'tails', 'tails', 'heads', 'tails', 'tails', 'heads',
       'heads', 'tails', 'tails', 'heads', 'tails', 'tails', 'tails',
       'tails', 'tails', 'tails', 'heads', 'tails', 'heads', 'tails',
       'heads', 'tails', 'tails', 'tails', 'heads', 'tails', 'tails',
       'heads', 'tails', 'heads', 'tails', 'heads', 'tails', 'tails',
       'tails', 'heads', 'tails', 'tails', 'heads', 'tails', 'heads',
       'tails', 'tails', 'tails', 'tails', 'tails', 'tails', 'heads',
       'heads', 'tails', 'heads', 'heads', 'tails', 'tails', 'tails',
       'tails', 'heads', 'tails', 'tails', 'heads', 'heads', 'heads',
       'heads', 'heads', 'heads', 'tails', 'heads', 'tails', 'heads',
       'heads', 'tails', 'tails', 'tails', 'heads', 'tails', 'tails',
       'heads', 'heads', 'heads', 'tails', 'tails', 'heads', 'tails',
       'tails', 'tails'], dtype='<U5')
```

[47]:
```
sum(tosses == 'heads')
```

[47]:
```
37
```

[48]:
```
outcomes = make_array()
```

[49]:
```
outcomes
```

[49]:
```
array([], dtype=float64)
```

[50]:
```
np.append(outcomes, sum(np.random.choice(coin, 100) == 'heads'))
```

[50]:
```
array([49.])
```

[51]:
```
outcomes
```

[51]:
```
array([], dtype=float64)
```

[52]:
```
outcomes = np.append(outcomes, sum(np.random.choice(coin, 100) == 'heads'))
outcomes
```

[52]:
```
array([45.])
```

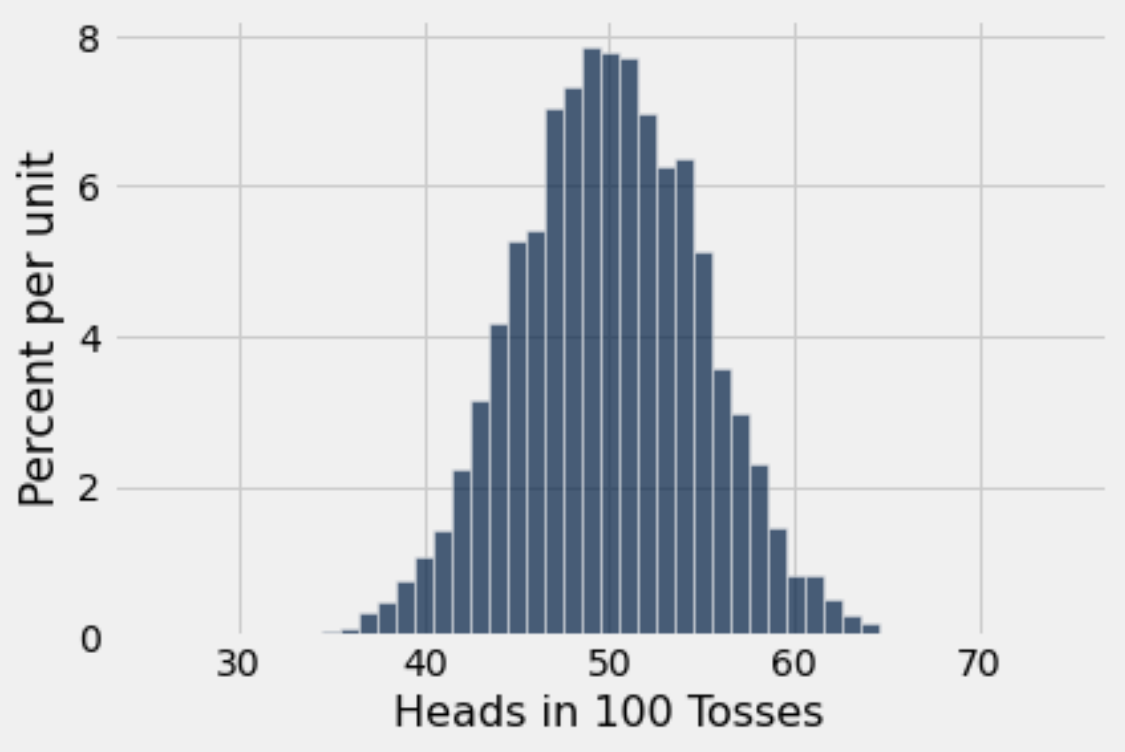
0.8 For Statements

[53]:
```
# the number of stations that have more than 500 starts and starts
outcomes = make_array()

for i in np.arange(7):
    outcomes = np.append(outcomes, sum(np.random.choice(coin, 100) == 'heads'))
outcomes
```

[53]:
```
array([46., 54., 62., 55., 45., 54., 52.])
```

[54]:
```
outcomes = make_array()

for i in np.arange(10000):
    outcomes = np.append(outcomes, sum(np.random.choice(coin, 100) == 'heads'))
```

[55]:
```
Table().with_column(
    'Heads in 100 Tosses', outcomes
).hist(bins = np.arange(25.5, 75.5))
```



[ ]:
```

```