

Boots



Bootstrap

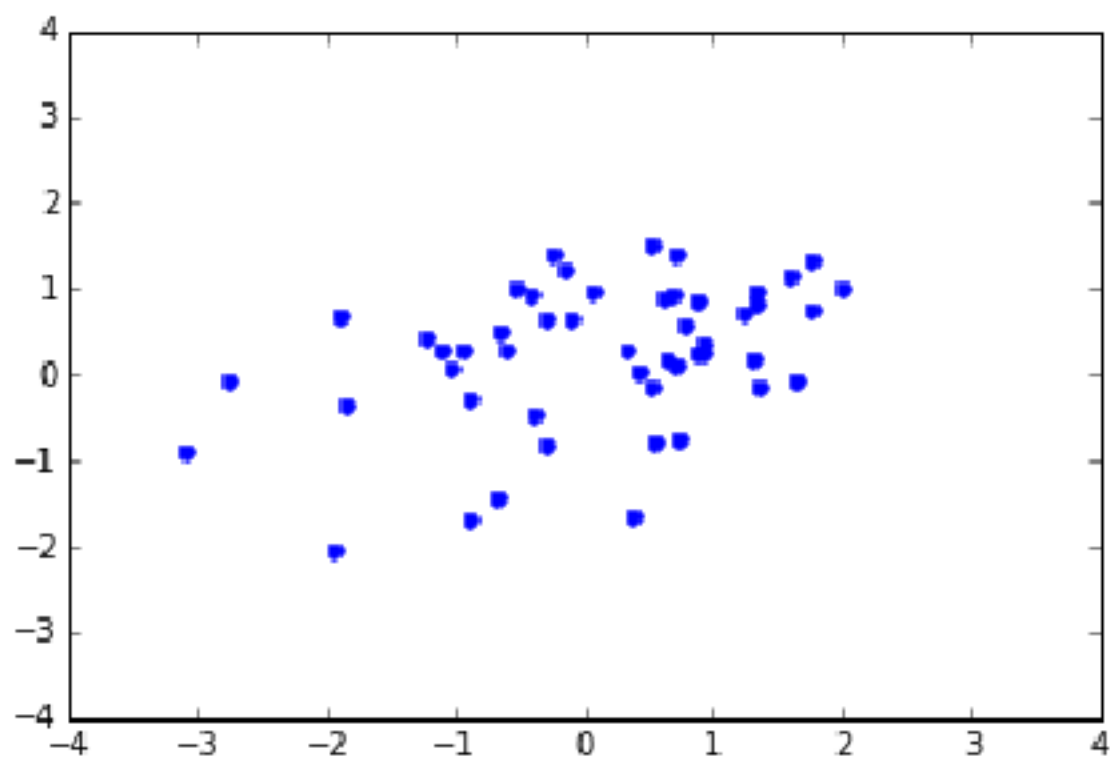
and phylogenetic trees



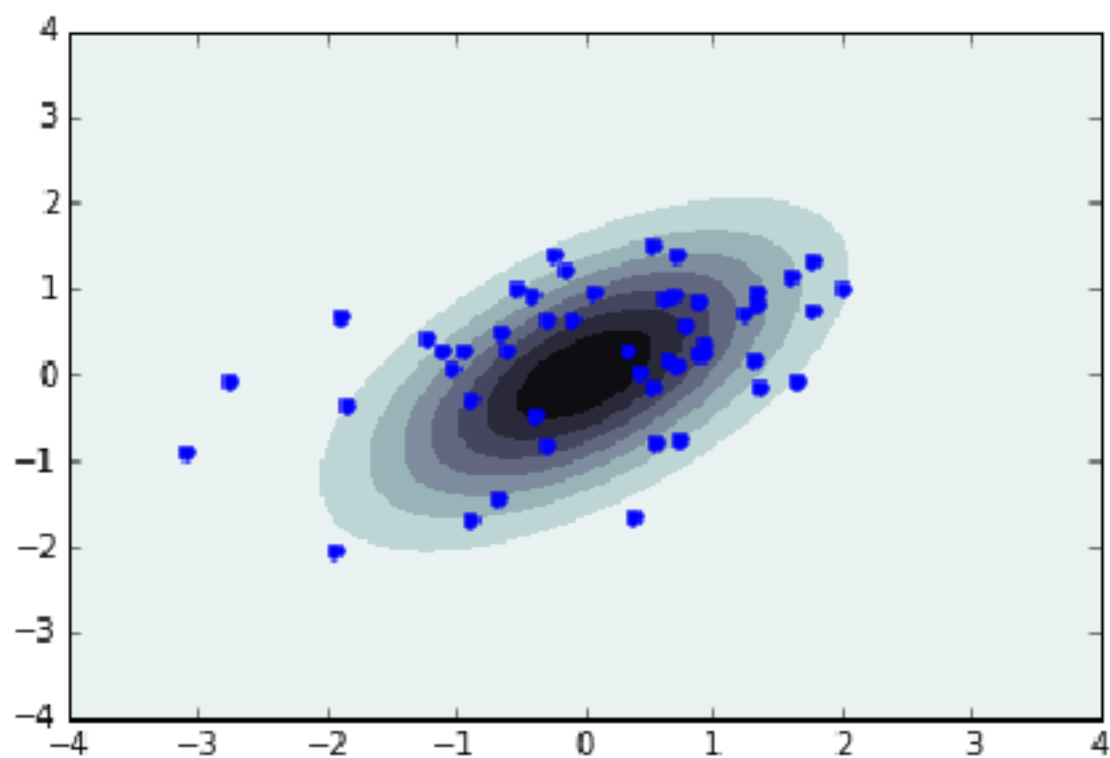
What does it mean

Tall boots may have a tab, loop or handle at the top known as a bootstrap, allowing one to use fingers or a boot hook tool to help pulling the boots on. The saying "to pull oneself up by one's bootstraps" was already in use during the 19th century as an example of an impossible task. The idiom dates at least to 1834, when it appeared in the Workingman's Advocate: "It is conjectured that Mr. Murphee will now be enabled to hand himself over the Cumberland river or a barn yard fence by the straps of his boots." In 1860 it appeared in a comment on metaphysical philosophy: "The attempt of the mind to analyze itself [is] an effort analogous to one who would lift himself by his own bootstraps." Bootstrap as a metaphor, meaning to better oneself by one's own unaided efforts, was in use in 1922.

Take a sample



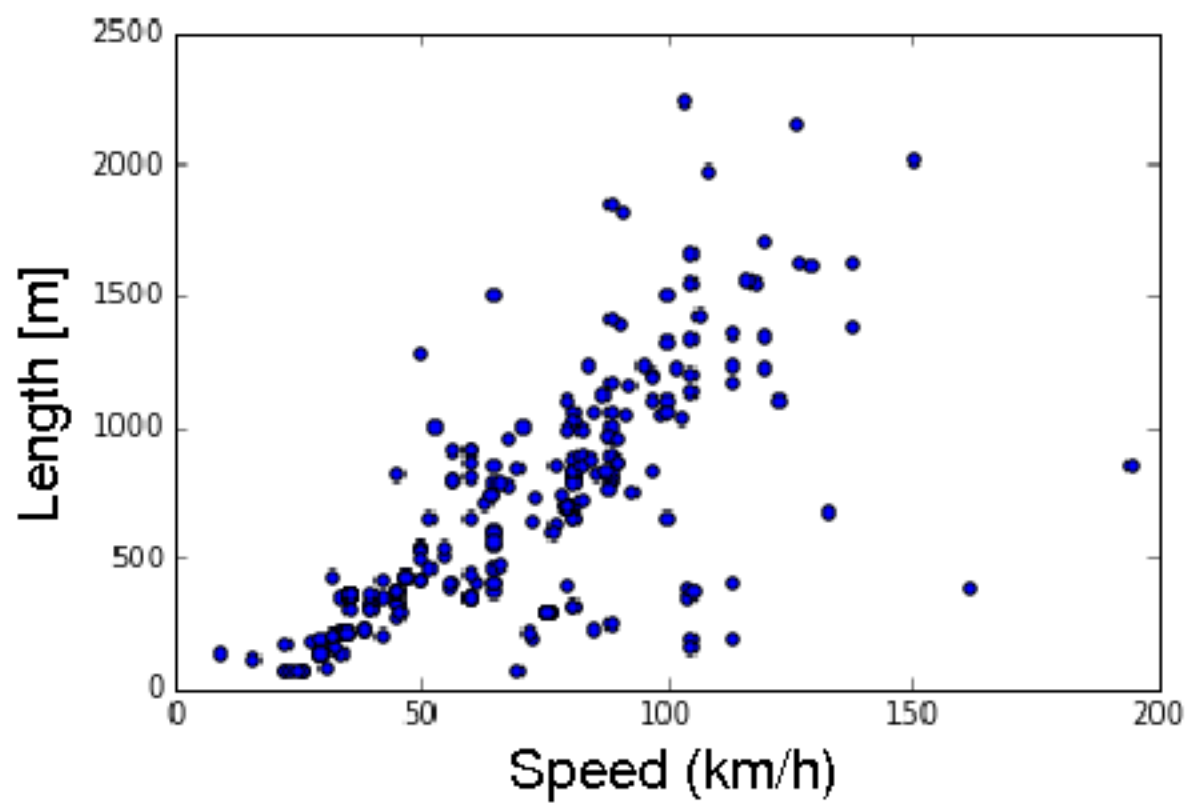
Take a sample



Roller Coaster

A photograph of a roller coaster with a red track and teal supports. The track is looping and curving against a blue sky with light clouds. The roller coaster is the central focus of the image.

Length and Speed of Roller coasters



Standard Analysis

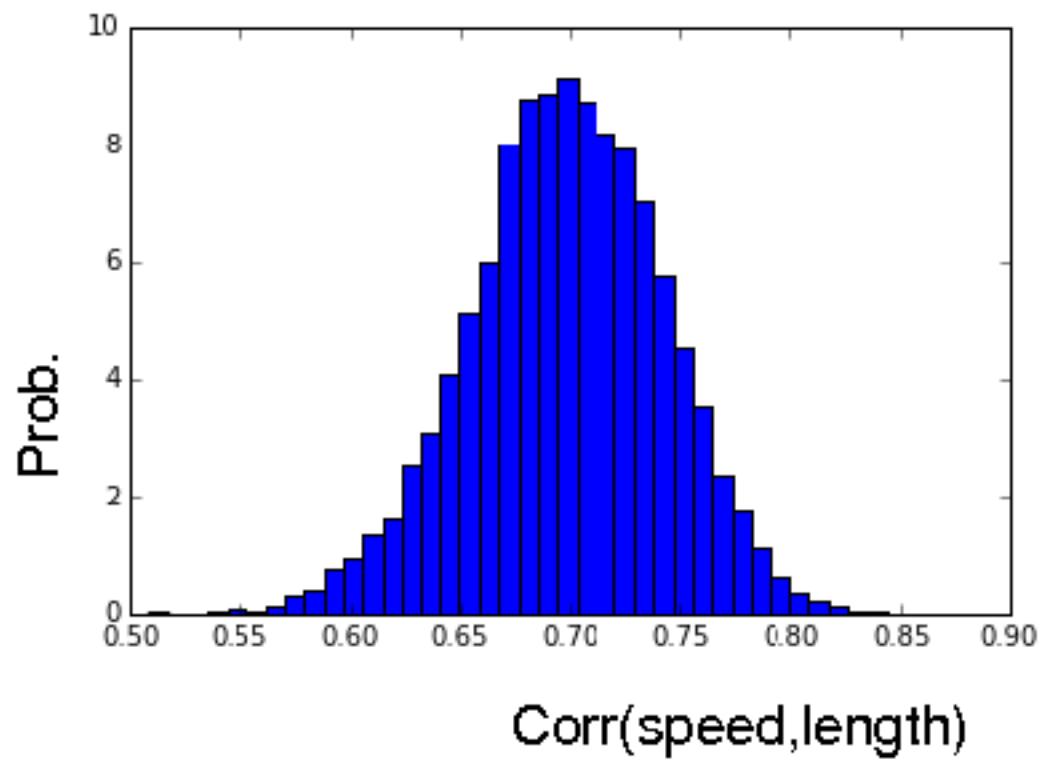
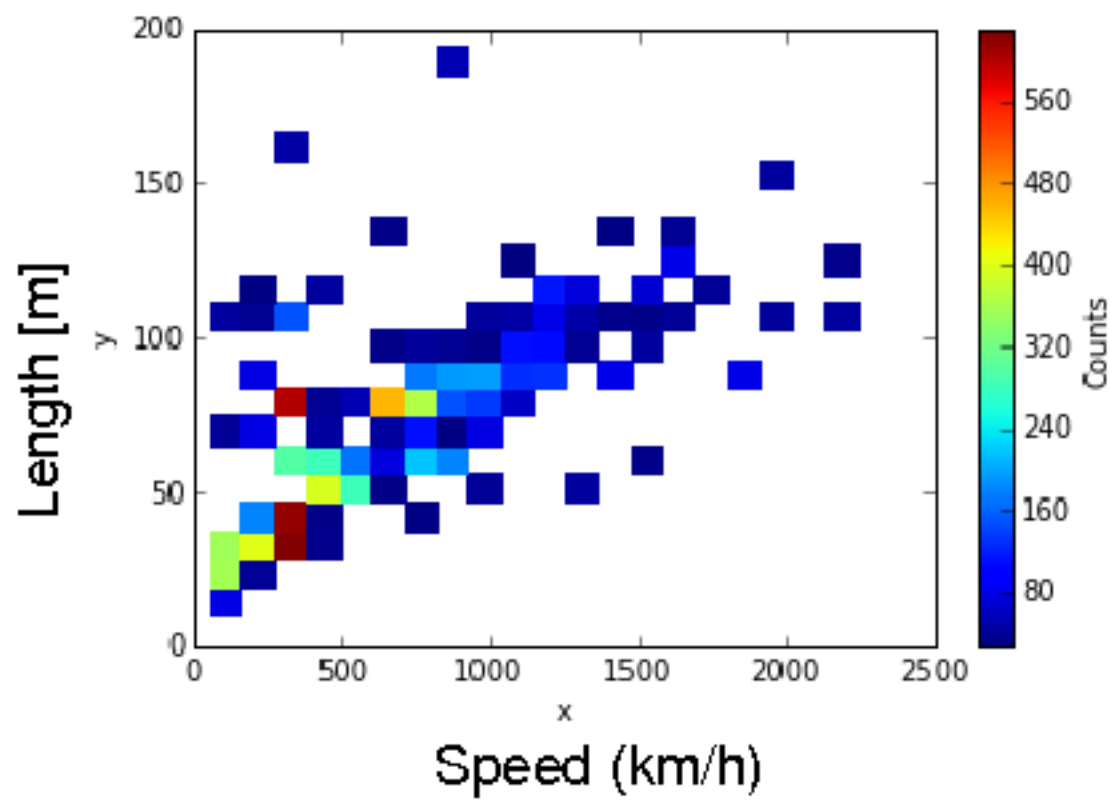
```
scipy.stats.describe(v)
```

```
DescribeResult(nobs=260, minmax=(array([ 9.72, 60. ]), array([ 194.4 ,  
2243.02])), mean=array([ 69.36838462, 648.00698077]), variance=array([  
865.4759495 , 196697.51442304]), skewness=array([ 0.56402121, 1.01759565  
]), kurtosis=array([ 0.56345422, 0.73611374]))
```

```
bpx,bpy = zip(*v); scipy.stats.pearsonr(bpx,bpy)
```

```
(0.69629308341240093, 4.9366122614103826e-39)
```


Bootstrap Analysis



Bootstrap Analysis

```
In [52]: pp=(0.69629308341240093, 4.9366122614103826e-39);z = np.arctanh(pp[0])
```

```
In [65]: sigma = (1/((260-3)**0.5));sigma
```

```
Out[65]: 0.06237828615518053
```

```
In [66]: cint = z + np.array([-1, 1]) * sigma * stats.norm.ppf((1+0.95)/2)
cint
```

```
Out[66]: array([ 0.73780954,  0.98232793])
```

```
In [67]: np.tanh(cint)
```

```
Out[67]: array([ 0.62781991,  0.75407188])
```

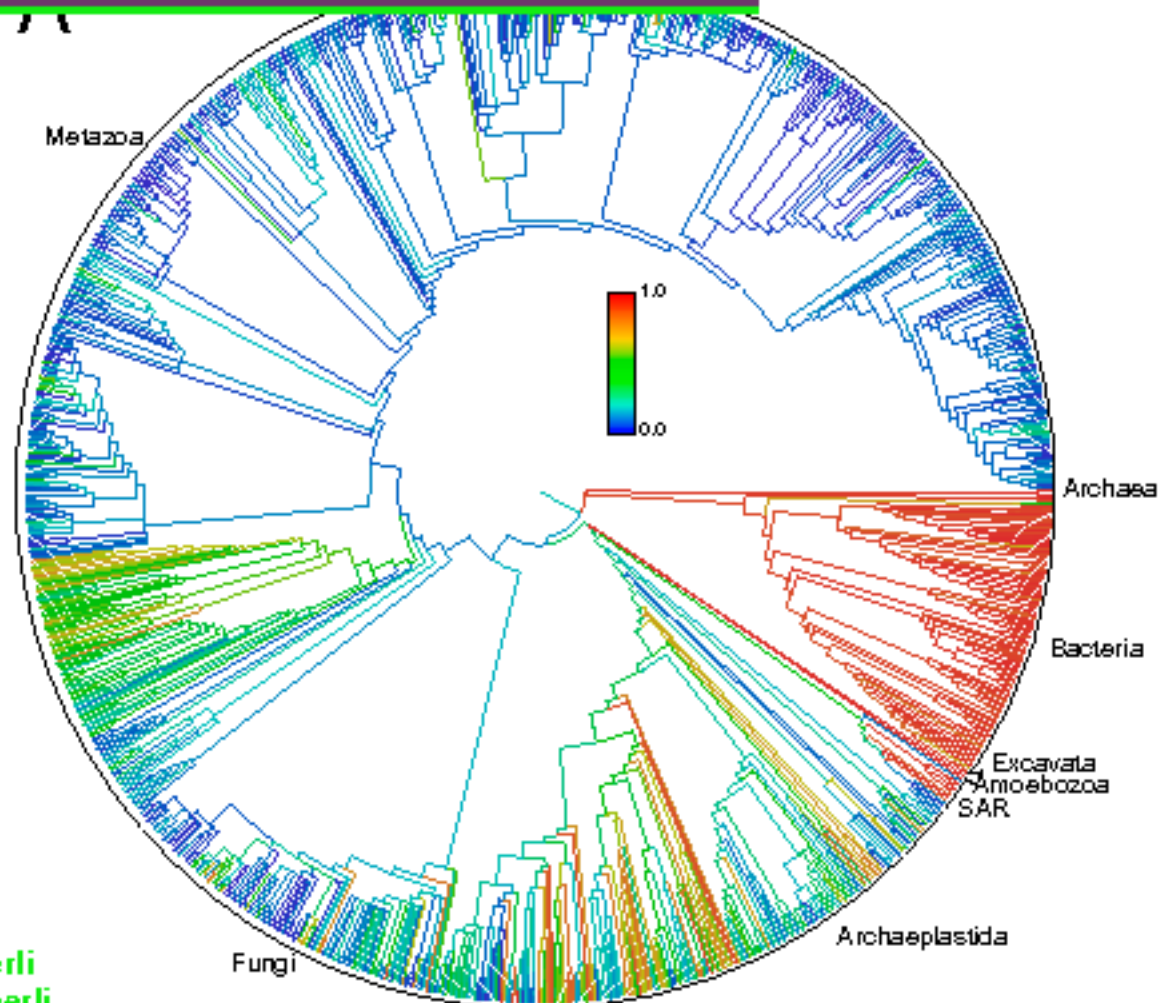
```
In [76]: scipy.cumsum(bins)/(scipy.sum(bins))
```

```
Out[76]: array([ 0.01811112,  0.03653619,  0.05527523,  0.07432821,  0.09369516,
                0.11337606,  0.13337091,  0.15367973,  0.1743025 ,  0.19523922,
                0.2164899 ,  0.23805454,  0.25993313,  0.28212568,  0.30463219,
                0.32745265,  0.35058707,  0.37403545,  0.39779778,  0.42187407,
                0.44626431,  0.47096851,  0.49598667,  0.52131878,  0.54696485,
                0.57292487,  0.59919885,  0.62578679,  0.65268869,  0.67990454,
                0.70743434,  0.73527811,  0.76343582,  0.7919075 ,  0.82069313,
                0.84979272,  0.87920626,  0.90893376,  0.93897522,  0.96933063,  1
                ])
```

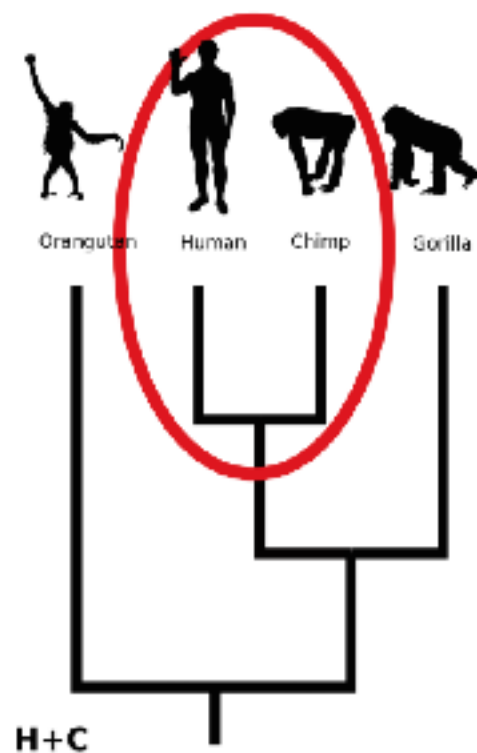
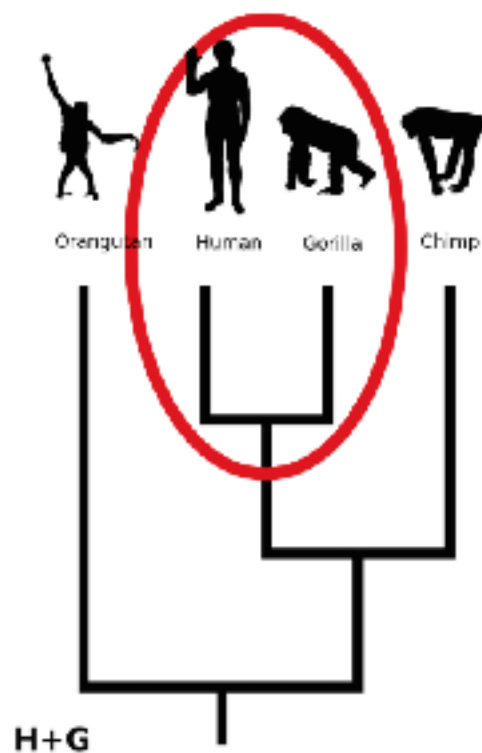
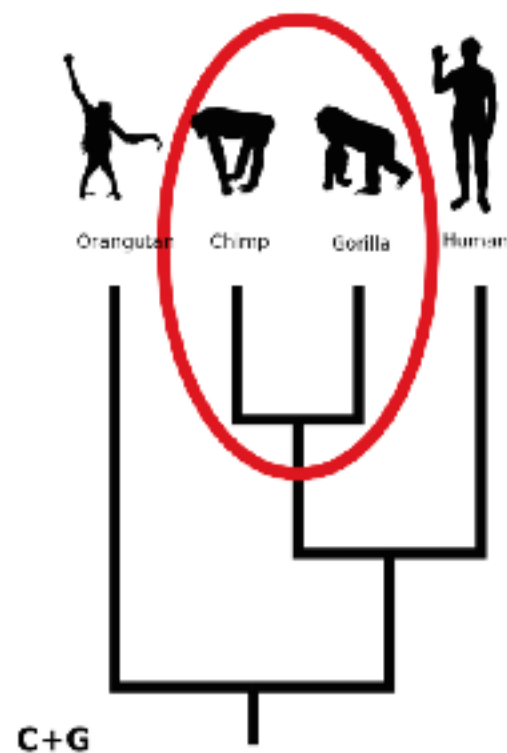
```
In [83]: (np.array(range(40))*0.01 + 0.5)[3], (np.array(range(40))*0.01 + 0.5)[-3],
```

```
Out[83]: (0.53000000000000003, 0.87)
```

Phylogenies



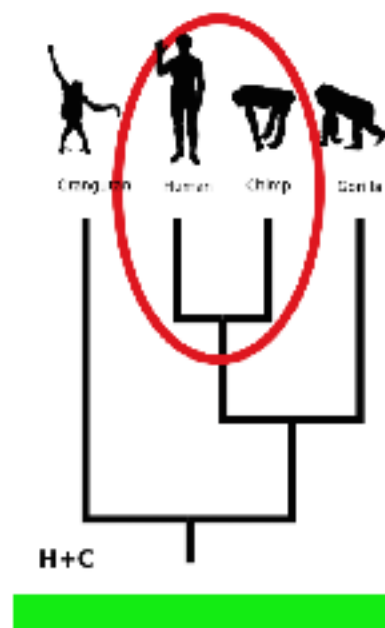
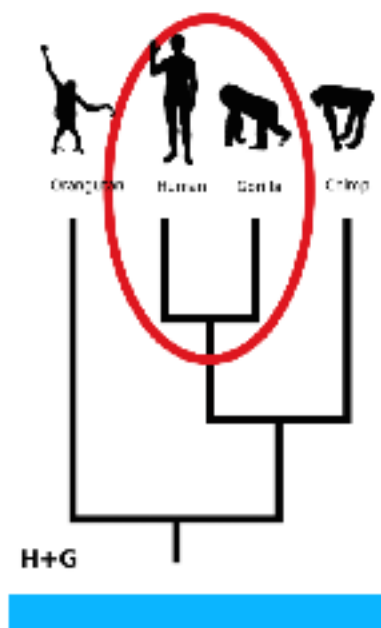
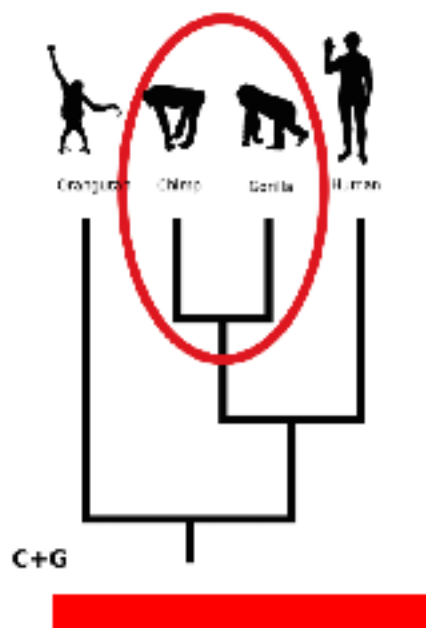
Bootstrapping phylogenies



by Mark Holder

Bootstrapping phylogenies

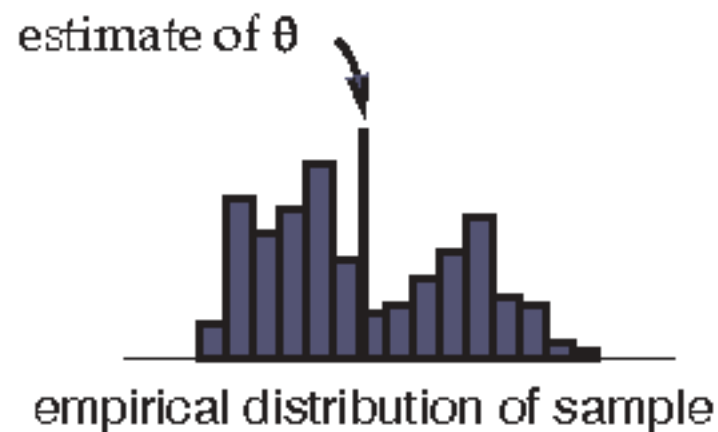
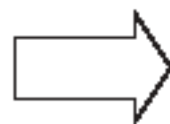
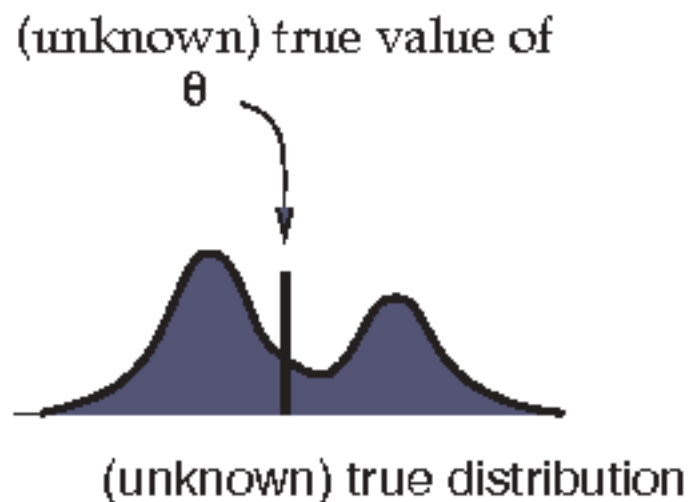
Human	G	C	C	A	G	C	T	G	C	T
Chimp	G	C	C	G	G	G	T	A	T	T
Gorilla	G	T	C	A	A	C	T	A	T	G
Orang	C	T	C	A	A	G	T	G	C	G



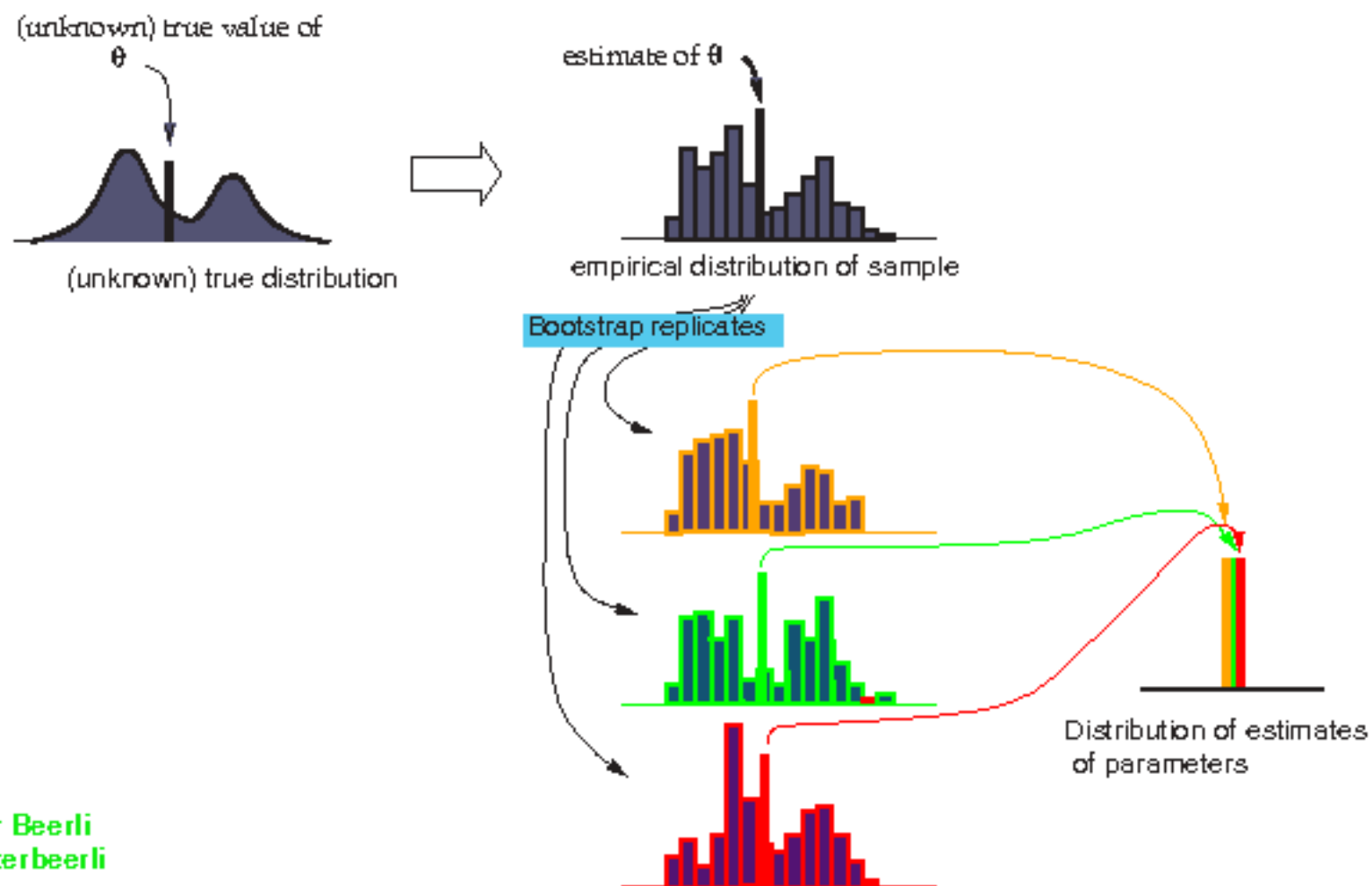
<http://phylo.bio.ku.edu/mephytis/boot-sample.html>

Bootstrapping phylogenies

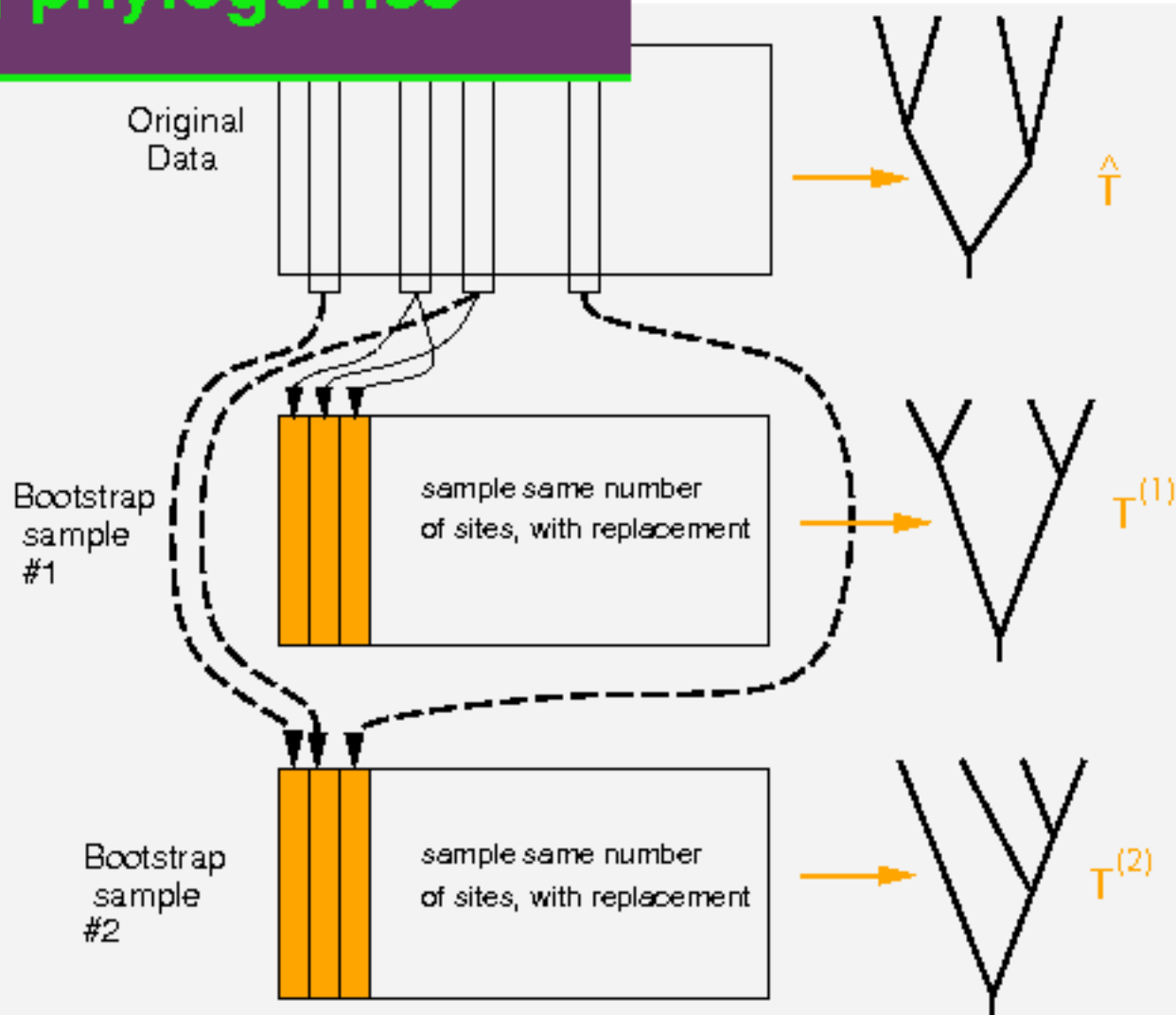
The bootstrap



Bootstrapping phylogenies



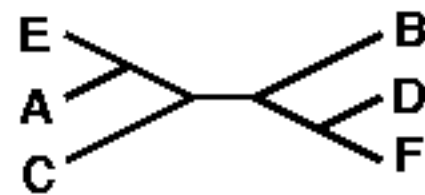
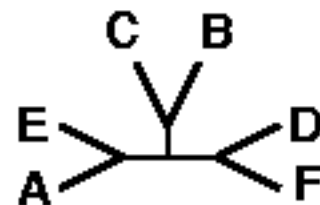
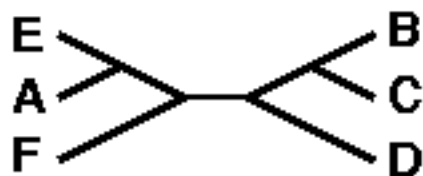
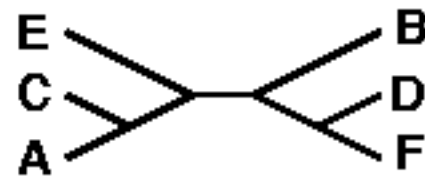
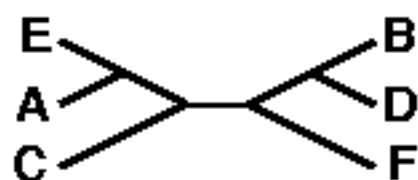
Bootstrapping phylogenies



Bootstrapping

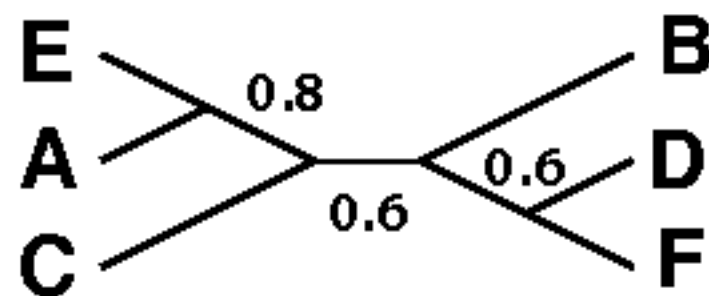
The majority-rule consensus tree

Trees:

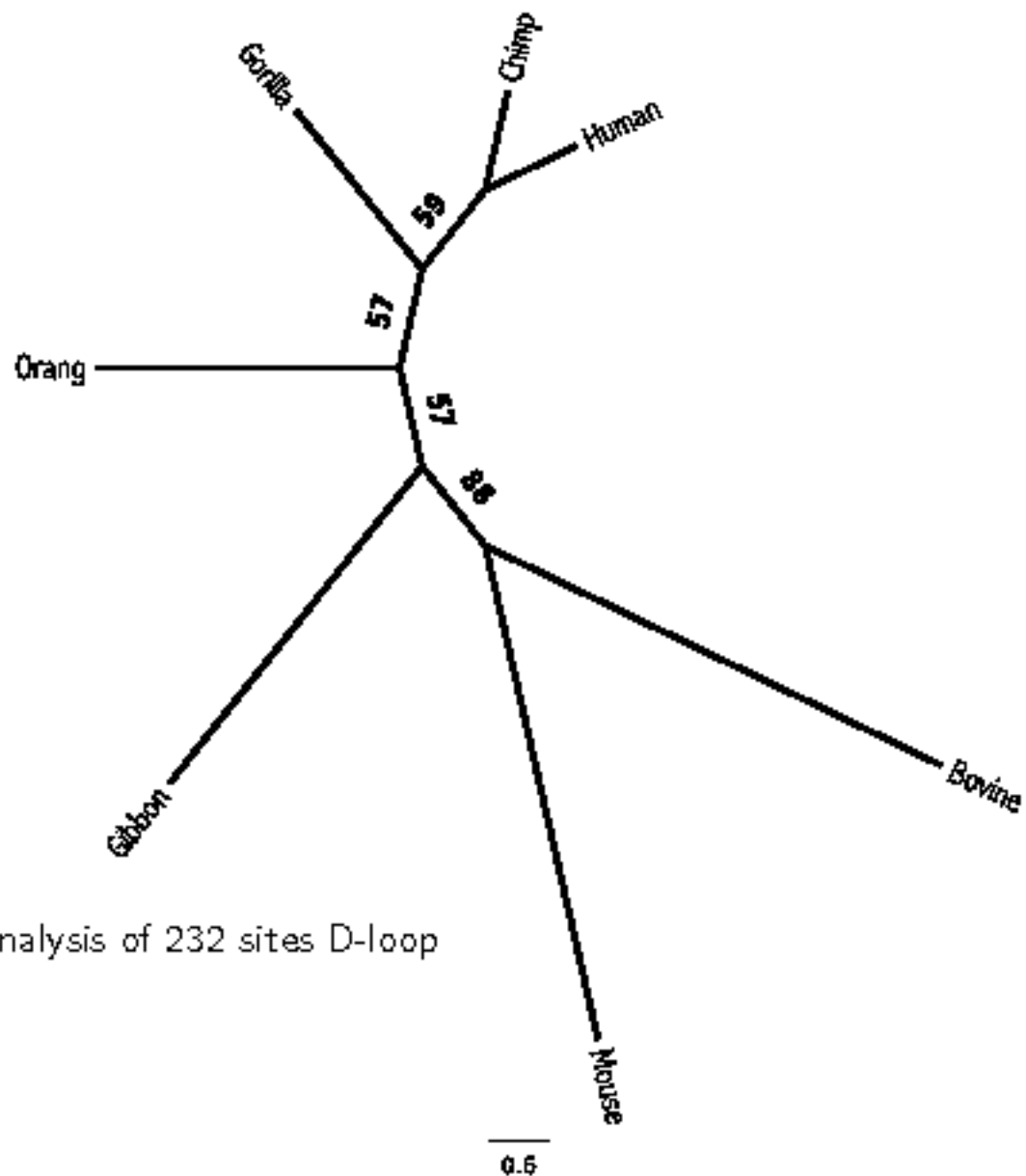


How many times each partition of species is found:

AE BCDF	4
ACE BDF	3
ACEF BD	1
AC BDEF	1
AEF BCD	1
ADEF BC	2
ABCE DF	3



Bootstrapping



From Hasegawa's analysis of 232 sites D-loop

Summary

The bootstrap allows you to generate a distribution based on your sample. This allows to take into account unknown correlation structure among the data entries.

The bootstrap has also problem in that we many need to block sample to consider correlations among sampling entries (for example sites in DNA sequences are correlated, any scheme that draws bootstrap samples independently for each site may be flawed.)

A great tool to handle sampling that deviates from normal distributions.