

Strava segment analysis

In[183]:=

```
In[184]:= raw = Import[ToFileName[NotebookDirectory[] <> "/data/", "segment1.csv"]];
```

```
In[185]:= colnames = Flatten[Take[raw, 1]]
```

Out[185]= {Date, Speed_Kmh, Power_W, Time_M_S}

```
In[186]:= data = Drop[raw, 1];
```

```
dates = Map[DateObject, data[[All, 1]]];
```

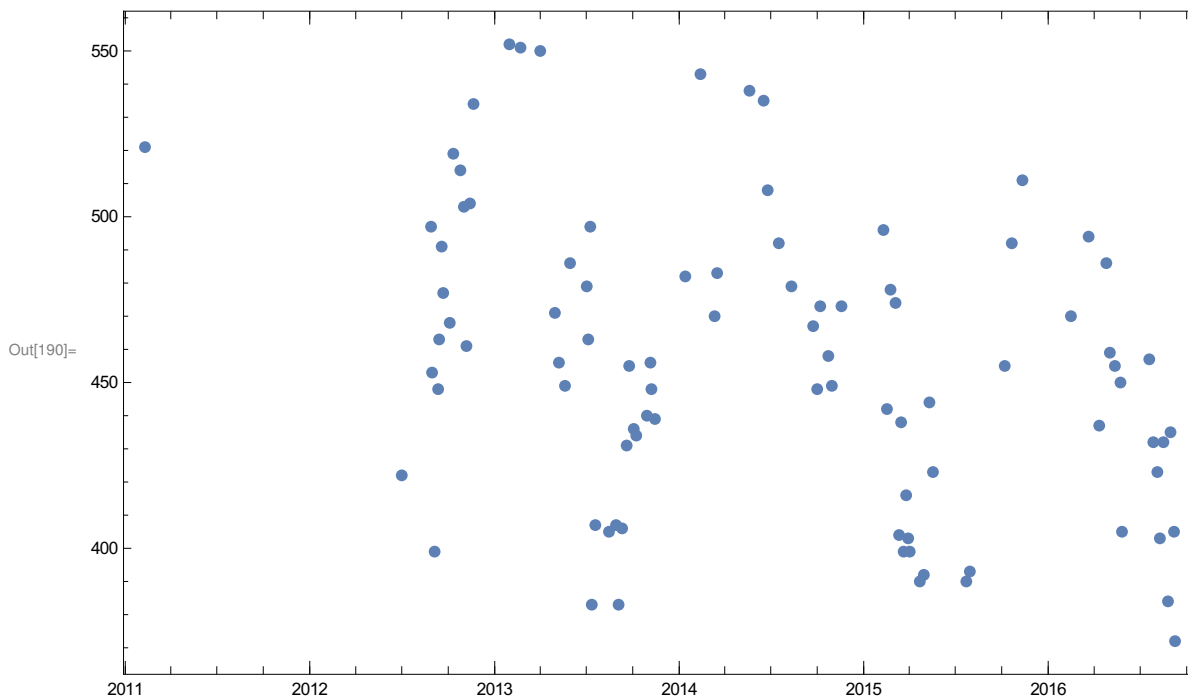
```
In[188]:= durations = Map[Function[s, With[{l = StringSplit[s, ":"]},  
    ToExpression[l[[1]] 60 + ToExpression[l[[2]]]]], data[[All, 4]]];
```

List of {date,duration} paris:

```
In[189]:= dd = Transpose[{dates, durations}];
```

Plot of all ride durations:

```
In[190]:= DateListPlot[dd, Joined -> False, ImageSize -> Large]
```



Let us try to find statistical distribution of all ride durations 2011-2016. We will consider only following distribution models to avoid more exotic ones:

```
In[191]:= dismodels = {UniformDistribution, NormalDistribution, GammaDistribution};
```

```
In[192]:= dist = FindDistribution[durations, TargetFunctions → dismodels]
```

```
Out[192]= MixtureDistribution[{0.386438, 0.613562},
  {NormalDistribution[453.878, 27.6987], UniformDistribution[{371.524, 551.804}]}
```

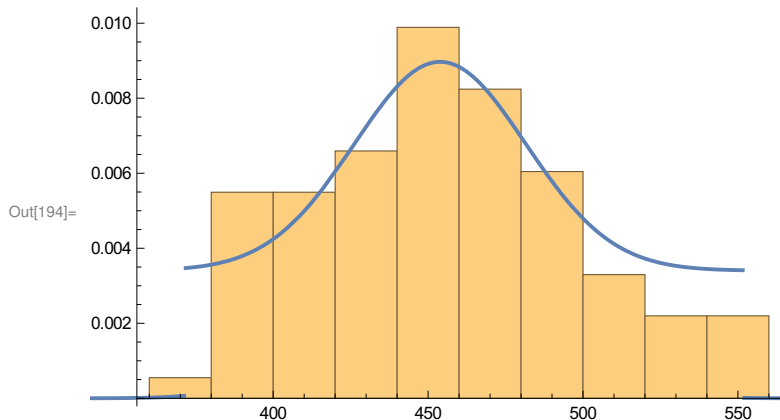
So the distribution is a mixture of uniform and normal distributions. Let us examine it more closely:

```
In[193]:= distpdf = PDF[dist, x]
```

```
Out[193]= { 0. + 0.00556584 e-0.000651707 (-453.878+x)2      x > 551.804 || x < 371.524
  0.00340339 + 0.00556584 e-0.000651707 (-453.878+x)2    True
```

Let us overlap estimated PDF on top of histogram:

```
In[194]:= Show[Histogram[durations, 10, "ProbabilityDensity", ImageSize → Medium],
  Plot[distpdf, {x, 84.318, 839.6}, ImageSize → Medium, PlotStyle → Thick]]
```



But we need to clear data little more:

1. I know that I started riding regularly in second half of 2012, so data before that is some kind of error
2. There is seems to be a seasonal component which could be explained that I usually ride more cautiously on wet road. We will ignore it for now.
3. Finally, it looks like the results change from year to year depending on my shape, so we will need to look at it by year.

Split by year and drop 2011:

```
In[195]:= ddbbyear = Drop[SplitBy[Sort[dd], DateList[#1[[1, 1]]][[1]] &, 1];
durationbyyear = Map[Last, ddbbyear, {2}];
```

Number of samples per year:

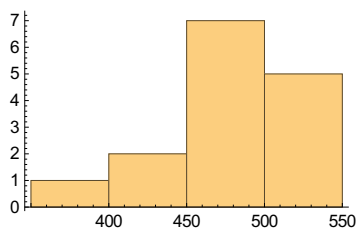
```
In[197]:= Map[Length, durationbyyear]
```

```
Out[197]= {15, 24, 15, 19, 17}
```

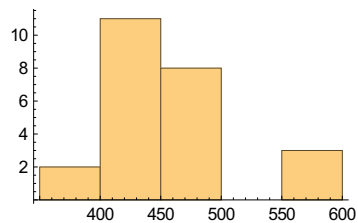
Let us see what distributions fit:

```
In[198]:= Map[{Histogram[#, ImageSize -> Small],  
FindDistribution[#, TargetFunctions -> dismodels]} &, durationbyyear] // TableForm
```

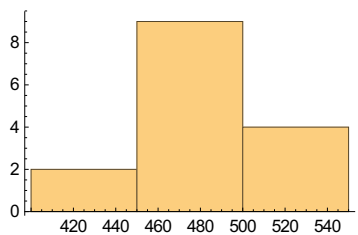
Out[198]//TableForm=



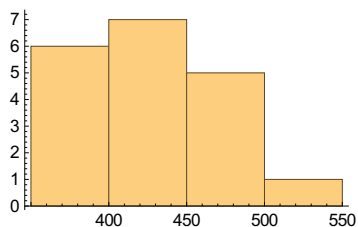
UniformDistribution[{398.699, 534.117}]



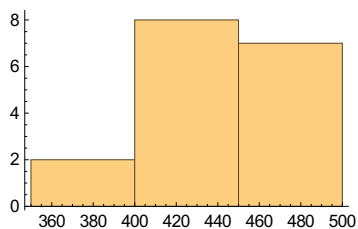
MixtureDistribution[{0.846207, 0.153793}, {NormalDistribution



MixtureDistribution[{0.76484, 0.23516}, {NormalDistribution



MixtureDistribution[{0.291808, 0.708192}, {NormalDistribution



UniformDistribution[{371.052, 493.488}]