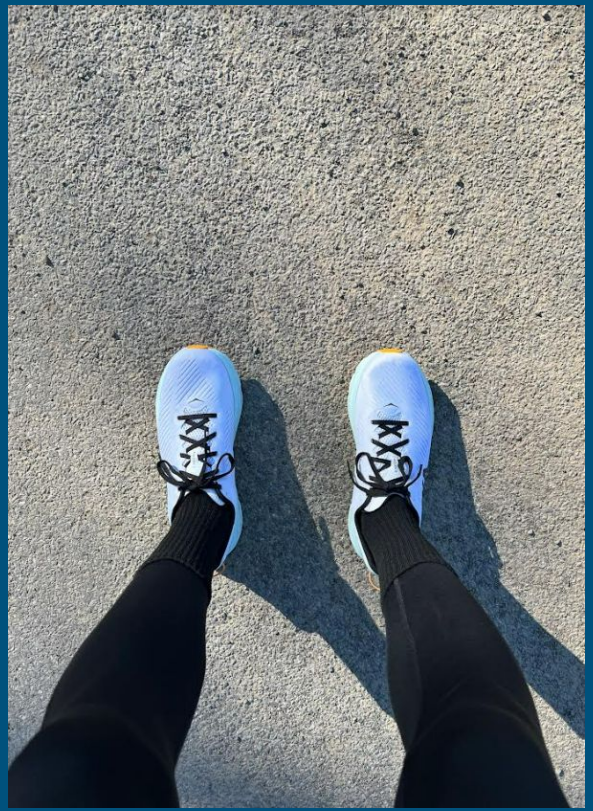
Take the Easy Route

Nicholas Scott
INFO-615-01
Spring 2022

Research Question

- As a runner, I end up doing a handful of routes over and over again, week in and week out
- I have some instincts and preconceptions about which routes are “easier” and which are “harder”, but I want to try and quantify those concepts and make a more scientific determination.
- Which running route is easiest?



Data

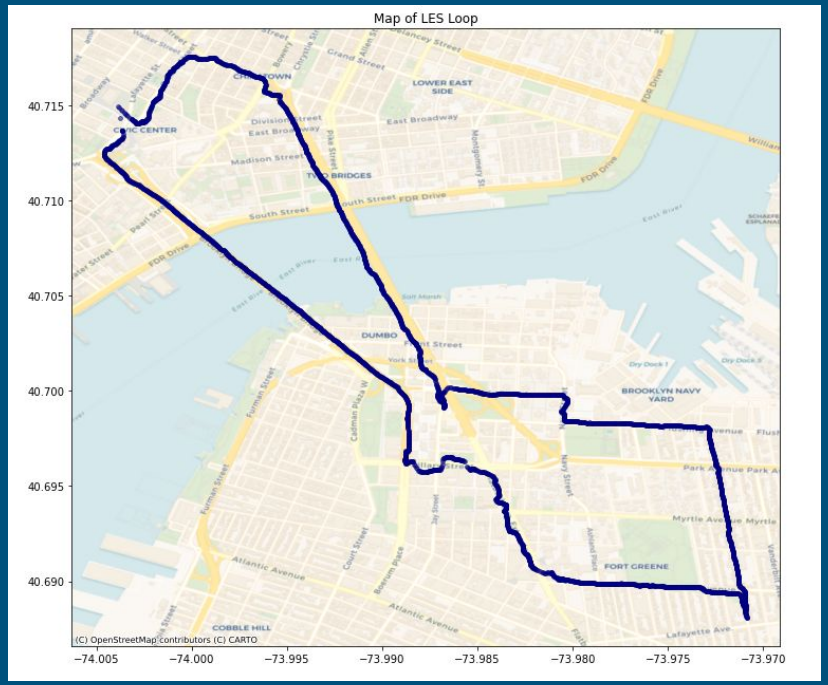
- I use a fitness tracking app called Strava, which captures, stores, and visualizes all of my runs.
- Strava allows users to download files from each activity in a .gpx format.
- I will select a few of these files to perform the analysis for this project.

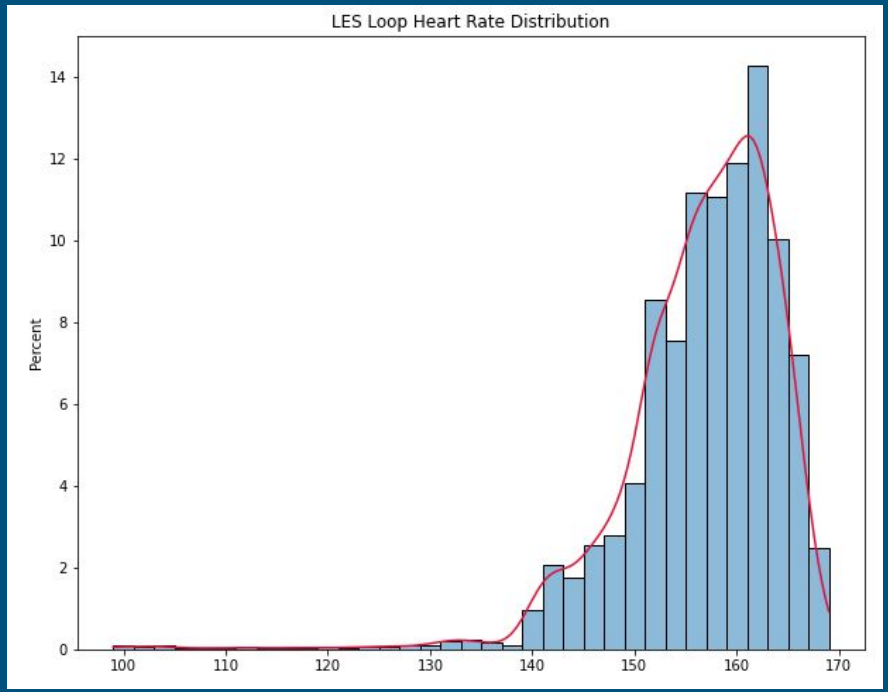
```
<?xml version="1.0" encoding="UTF-8"?>
<gpx creator="StravaGPX" xmlns:xsi="http://www.w3.org/2001/XML
http://www.garmin.com/xmlschemas/GpxExtensionsv3.xsd http://ww
xmlns:gpxtpx="http://www.garmin.com/xmlschemas/TrackPointExten
 <metadata>
  <time>2022-03-03T11:58:46Z</time>
 </metadata>
 <trk>
  <name>Everything hurts</name>
  <type>9</type>
  <trkseg>
   <trkpt lat="40.6884970" lon="-73.9708980">
    <ele>25.5</ele>
    <time>2022-03-03T11:58:46Z</time>
    <extensions>
     <gpxtpx:TrackPointExtension>
      <gpxtpx:hr>103</gpxtpx:hr>
      <gpxtpx:cad>83</gpxtpx:cad>
     </gpxtpx:TrackPointExtension>
    </extensions>
   </trkpt>
   <trkpt lat="40.6885140" lon="-73.9708980">
    <ele>25.4</ele>
    <time>2022-03-03T11:58:51Z</time>
    <extensions>
     <gpxtpx:TrackPointExtension>
      <gpxtpx:hr>99</gpxtpx:hr>
      <gpxtpx:cad>0</gpxtpx:cad>
     </gpxtpx:TrackPointExtension>
    </extensions>
   </trkpt>
   <trkpt lat="40.6885180" lon="-73.9709090">
    <ele>25.4</ele>
    <time>2022-03-03T11:58:52Z</time>
    <extensions>
     <gpxtpx:TrackPointExtension>
      <gpxtpx:hr>99</gpxtpx:hr>
      <gpxtpx:cad>0</gpxtpx:cad>
     </gpxtpx:TrackPointExtension>
    </extensions>
   </trkpt>
   <trkpt lat="40.6885310" lon="-73.9709220">
    <ele>25.4</ele>
    <time>2022-03-03T11:58:53Z</time>
    <extensions>
     <gpxtpx:TrackPointExtension>
      <gpxtpx:hr>100</gpxtpx:hr>
      <gpxtpx:cad>0</gpxtpx:cad>
     </gpxtpx:TrackPointExtension>
    </extensions>
   </trkpt>
```

EDAV

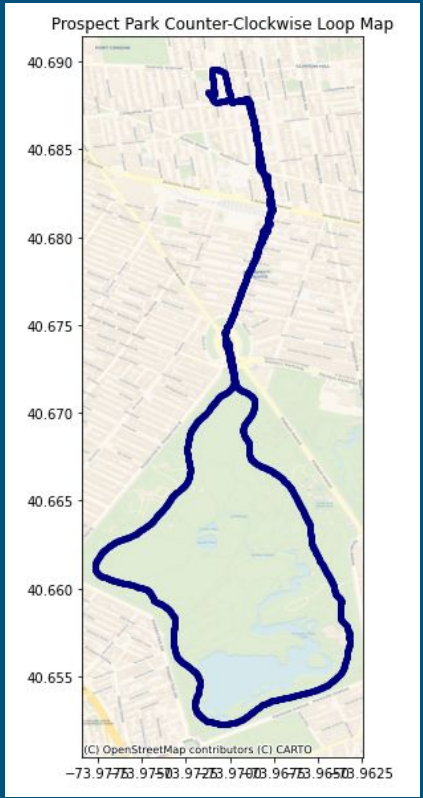
- I began by first mapping out each of the routes I’d selected, and then creating a histogram of the heart rates from each trackpoint.

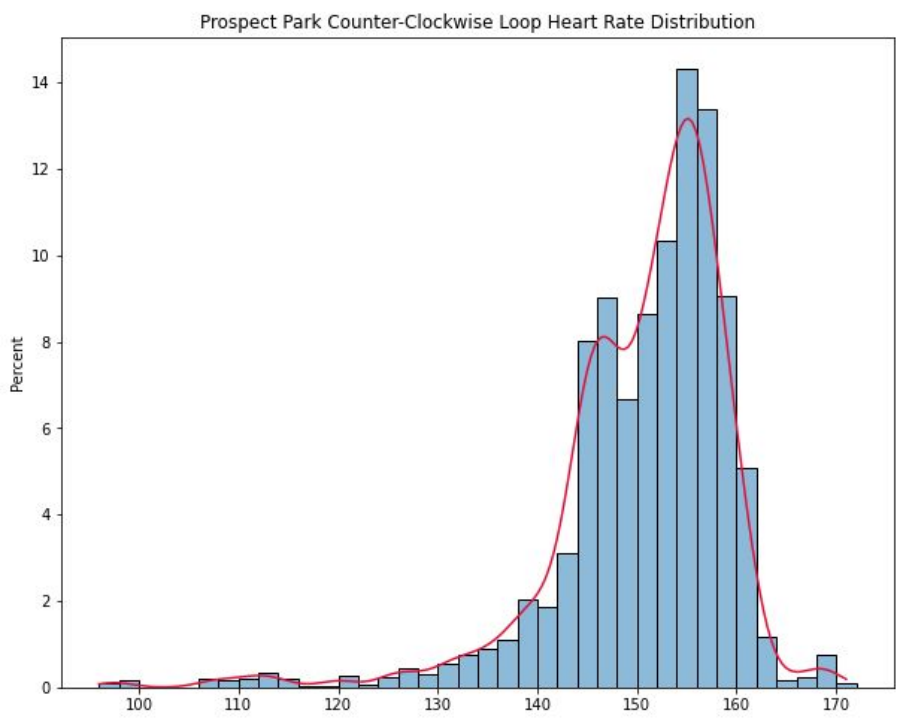
Lower East Side Loop



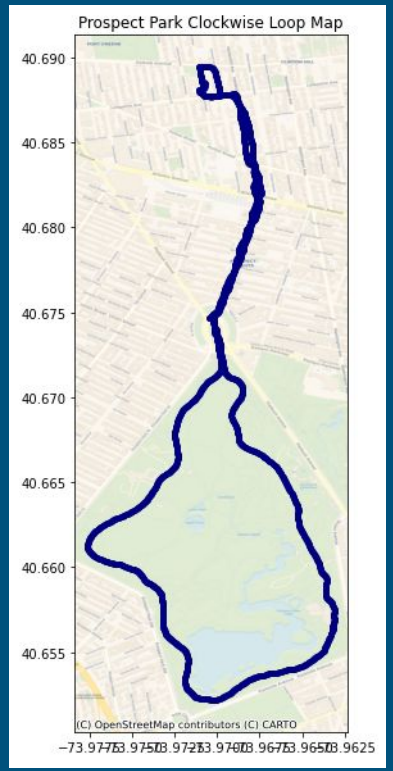


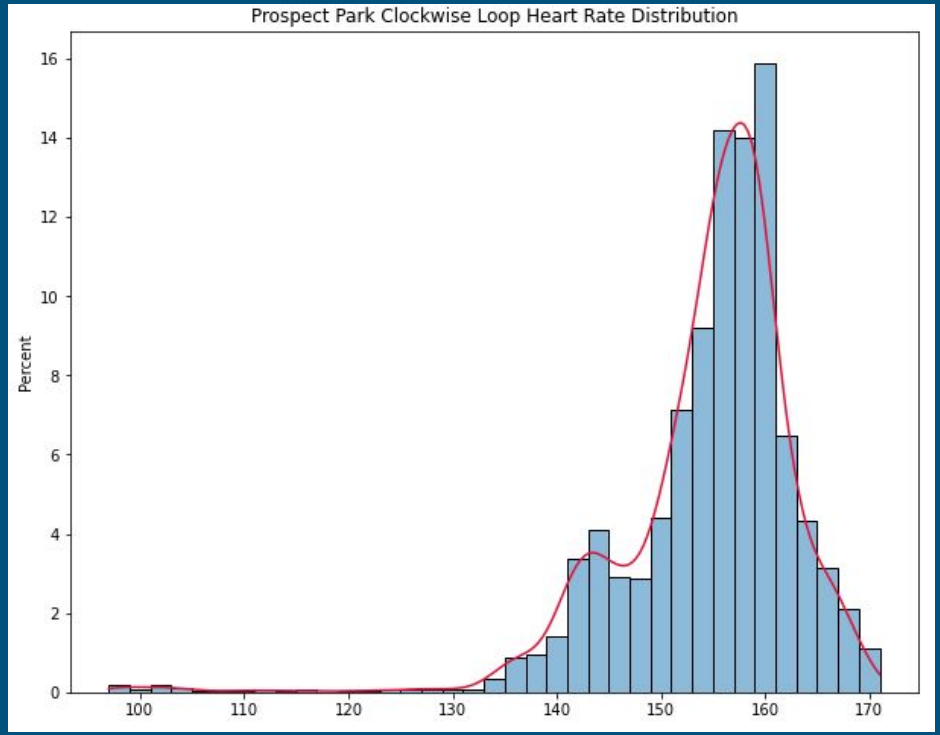
Prospect Park: Counter-Clockwise





Prospect Park: Clockwise





Model/Analysis

- I began the analysis by storing the heart rate data for each activity as a numpy array
- I then calculated the mean HR of each activity as a starting point to create a hypothesis about which route was “easiest”

```
LES_hr = LESloop_df['heart_rate'].to_numpy()

PPccw_hr = PPccw_df['heart_rate'].to_numpy()

PPcw_hr = PPcw_df['heart_rate'].to_numpy()
```

```
LES_hr.mean()
```

```
156.53628808864266
```

```
PPccw_hr.mean()
```

```
150.5704945992041
```

```
PPcw_hr.mean()
```

```
154.50350058343057
```

Check for Normalcy

- I checked each array for normalcy, and determined that none of them were normally distributed. My conclusion was that a Mann-Whitney U Test would be an appropriate method to compare each route.

```
stats.normaltest(LES_hr)
```

```
NormaltestResult(statistic=1687.053646845702, pvalue=0.0)
```

```
stats.normaltest(PPccw_hr)
```

```
NormaltestResult(statistic=1545.0429760179593, pvalue=0.0)
```

```
stats.normaltest(PPcw_hr)
```

```
NormaltestResult(statistic=1668.4523293705388, pvalue=0.0)
```

LES Loop vs Prospect Park Counter-Clockwise

- Null hypothesis: Both routes are identical in terms of effort

- Alternative hypothesis: The PP: CCW loop is easier

```
stats.mannwhitneyu(PPccw_hr, LES_hr, alternative='less')
```

```
MannwhitneyuResult(statistic=3422724.0, pvalue=9.441089337369227e-250)
```

Prospect Park Counter-Clockwise vs Clockwise

- Null hypothesis: Both routes are identical in terms of effort

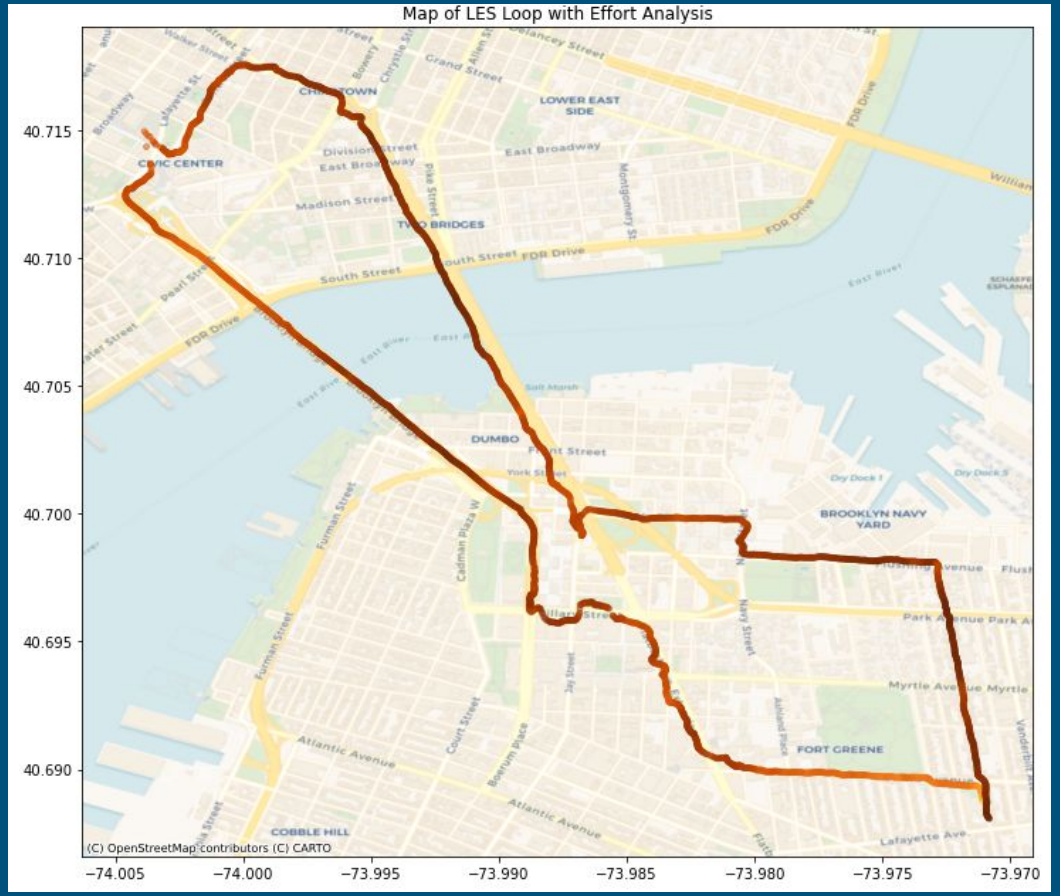
```
stats.mannwhitneyu(PPccw_hr, PPcw_hr, alternative='less')
```

```
MannwhitneyuResult(statistic=4120046.0, pvalue=3.371749048162538e-116)
```

- Alternative hypothesis: The PP: CCW loop is easier

Final Visualization

I wanted to visualize a route and incorporate heart rate data to indicate the “easy” parts and “hard” parts of the route



Feature Engineering

I created a feature to classify each trackpoint as “easy”, “medium”, or “hard” based on the heart rate

```
hr_mean = LES_hr.mean()
```

```
hr_std = LES_hr.std()
```

```
def create_effort_zones(x):
    # "easy" effort will be defined as a heart rate that is one standard deviation below the mean
    if x < (hr_mean - hr_std):
        return "1"
    # "medium" effort will be defined as a heart rate that is within one standard deviation above or below the mean
    elif x >= (hr_mean - hr_std) and x < (hr_mean + hr_std):
        return "2"
    # "hard" effort will be defined as a heart rate that is one standard deviation above the mean
    elif x >= (hr_mean + hr_std):
        return "3"
```

```
LESloop_df['effort_zone'] = LESloop_df['heart_rate'].apply(lambda x: create_effort_zones(x))
```

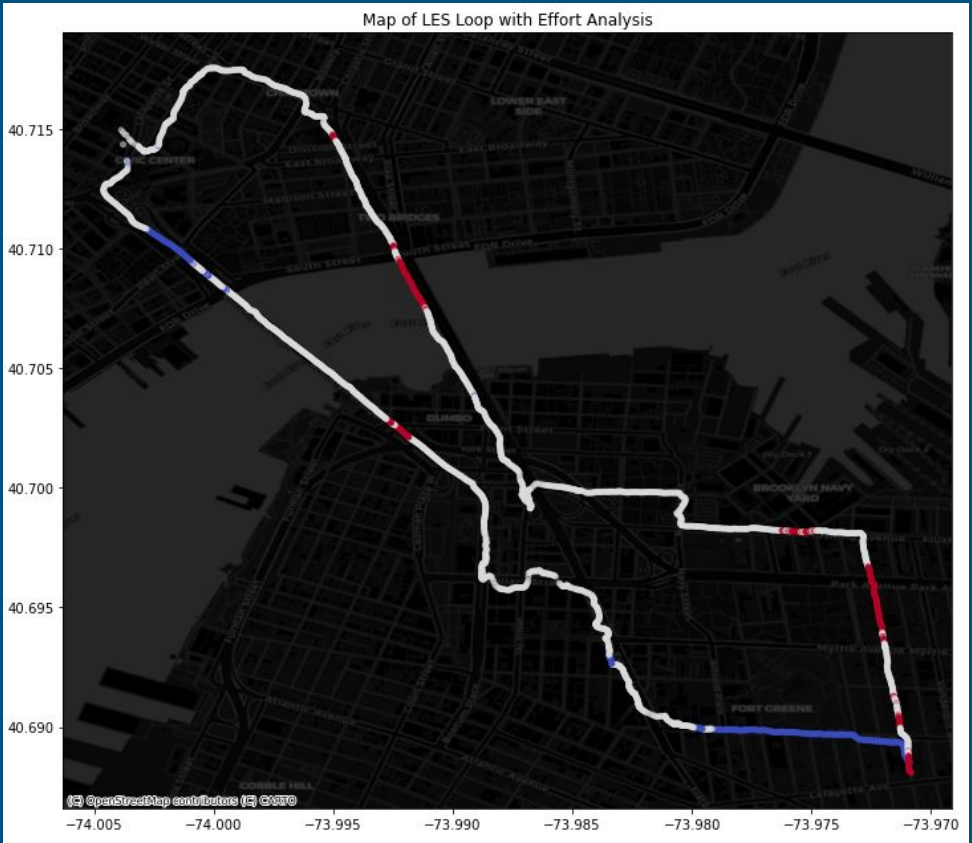
Feature Engineering Continued

LESloop_df

	Unnamed: 0	latitude	longitude	elevation	time	heart_rate	cadence	geometry	effort_zone
0	0	40.688497	-73.970898	25.5	2022-03-03 11:58:46+00:00	103	83	POINT (-73.970898 40.688497)	1
1	1	40.688514	-73.970898	25.4	2022-03-03 11:58:51+00:00	99	0	POINT (-73.970898 40.688514)	1
2	2	40.688518	-73.970909	25.4	2022-03-03 11:58:52+00:00	99	0	POINT (-73.970909 40.688518)	1
3	3	40.688531	-73.970922	25.4	2022-03-03 11:58:53+00:00	100	0	POINT (-73.970922 40.688531)	1
4	4	40.688545	-73.970934	25.4	2022-03-03 11:58:54+00:00	101	0	POINT (-73.970934 40.688545)	1
...	...	...	...	...	...	...	...	...	...
3605	3605	40.688159	-73.970896	26.2	2022-03-03 12:59:14+00:00	168	86	POINT (-73.970896 40.688159)	3
3606	3606	40.688134	-73.970880	26.3	2022-03-03 12:59:15+00:00	168	86	POINT (-73.97088 40.688134)	3
3607	3607	40.688112	-73.970860	26.3	2022-03-03 12:59:16+00:00	167	85	POINT (-73.97086 40.688112)	3
3608	3608	40.688089	-73.970846	26.4	2022-03-03 12:59:17+00:00	167	85	POINT (-73.970846 40.688089)	3
3609	3609	40.688073	-73.970829	26.4	2022-03-03 12:59:18+00:00	167	86	POINT (-73.970829 40.688073)	3

3610 rows × 9 columns

Final Visualization



The End

- All code and data for the project can be found on GitHub: https://github.com/nickxscott/TakeTheEasyRoute
- Thank you!