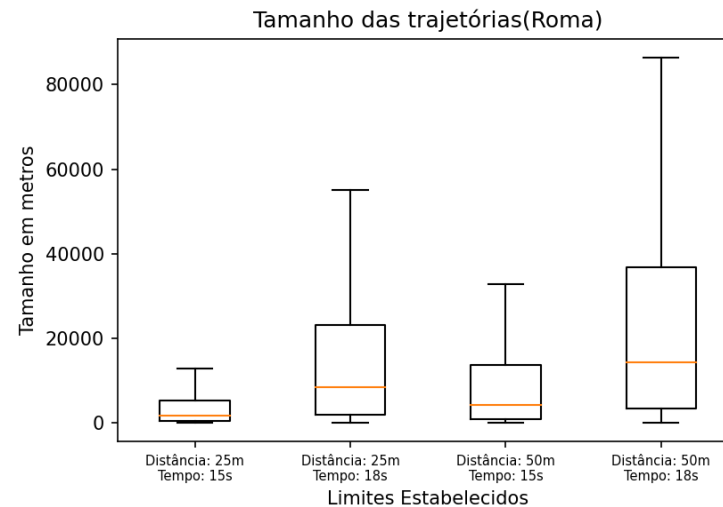
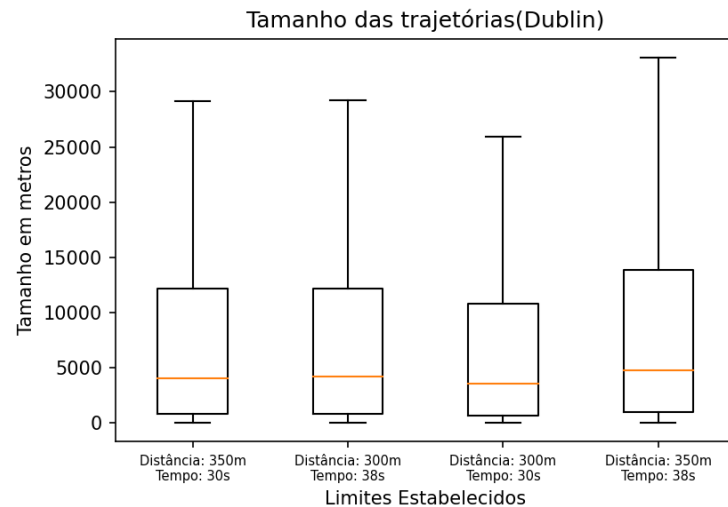


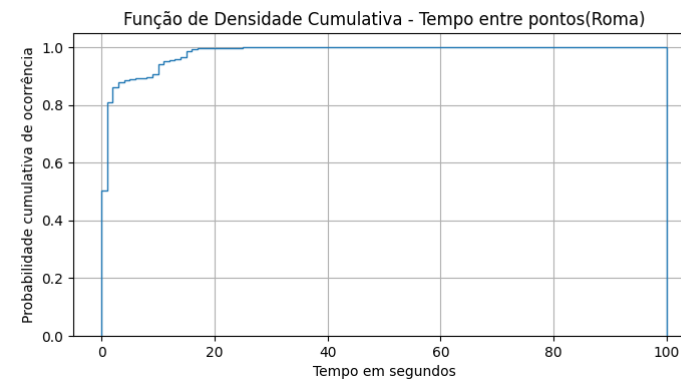
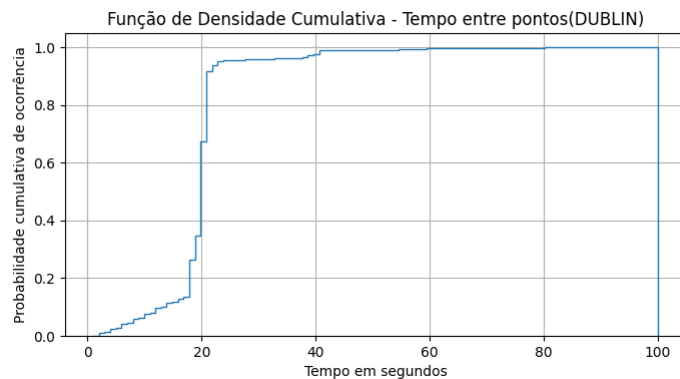
# Boxplots



Nos gráficos de boxplot, cada item foi construído tendo como base dois limites, de distância e tempo entre dois pontos, que foram escolhidos pela observação das CDFs de tempos e distâncias entre pontos. Analisando os boxplots de Dublin aos de Roma, vemos que a alteração de limites não interfere muito no tamanho das trajetórias. Já os de Roma sofrem uma grande mudança, principalmente pela alteração do limite de tempo. Entre os dois gráficos, vemos que as trajetórias de Roma são, em geral, maiores que as de Dublin se desconsiderarmos a primeira box. Isso pode ocorrer por conta da diferença de comportamento entre ônibus e táxis, bem como pela diferença estrutural das cidades, indicando que a cidade de Roma pode favorecer viagens mais longas.

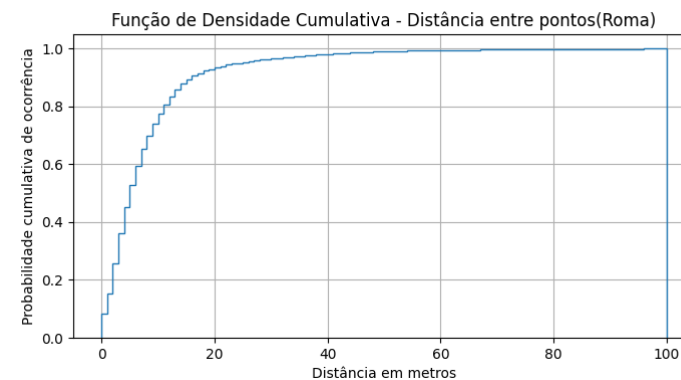
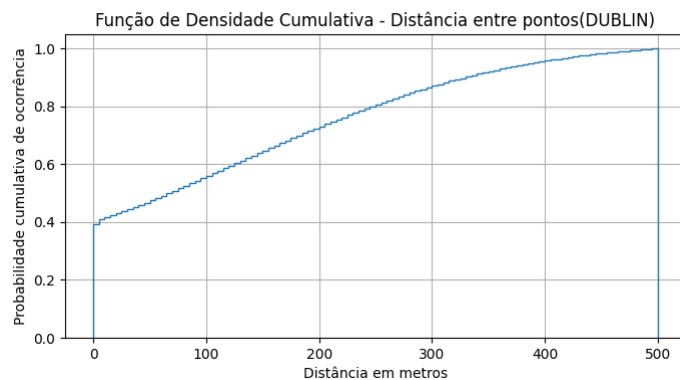
## Funções de Densidade Cumulativa

### Tempo entre pontos



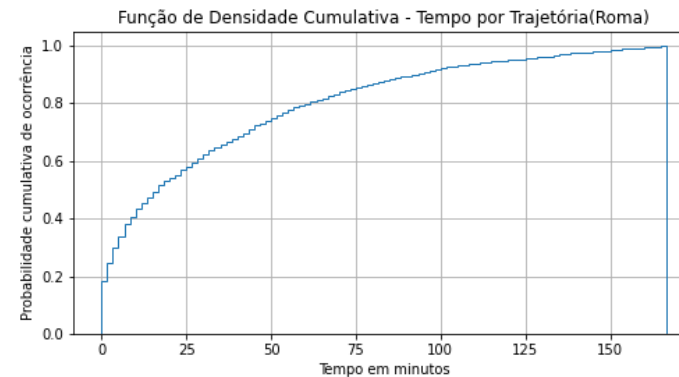
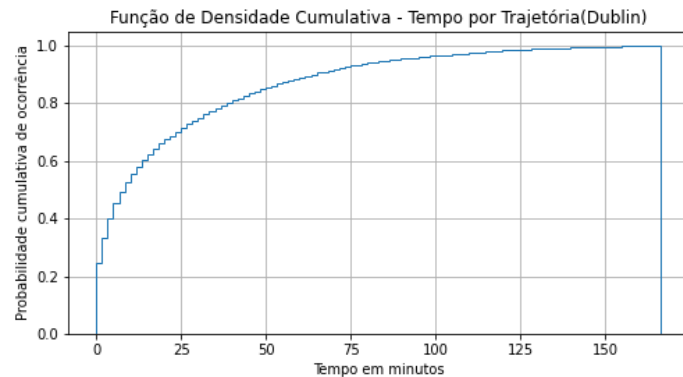
As CDFs de tempo entre dois pontos indicam a porcentagem de dados que está presente até determinado limite de tempo. Elas foram usadas para estabelecer o limite de tempo a ser utilizado no algoritmo de separação de viagens. No caso de Dublin os limites são 30 e 38 segundos, representando valores até aproximadamente 95% da base de dados. Em Roma, esses limites são 15 e 18 segundos, representando 95% e 98%, respectivamente.

## Distância entre pontos



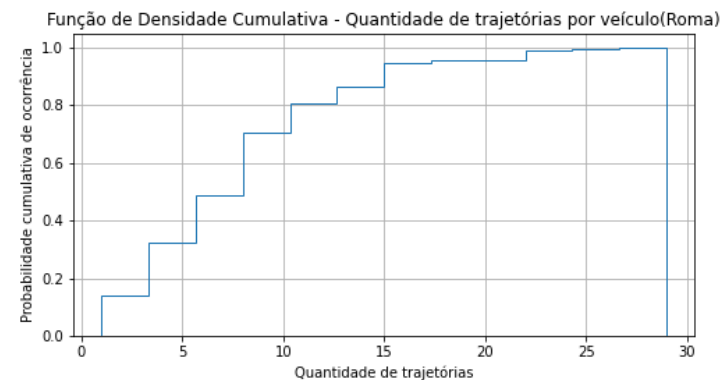
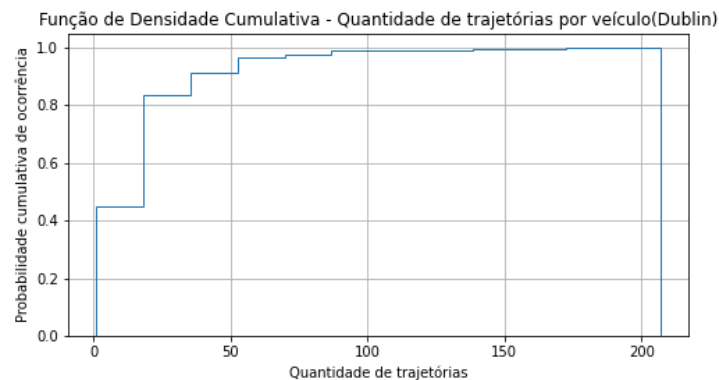
Como as CDFs anteriores, essas também foram utilizadas para estabelecer limites, neste caso de distância. Percebemos que a curva de Dublin é bem mais suave se comparada à de Roma, que mostra uma concentração grande de dados com valores pequenos. Em Roma, foram escolhidos os limites de 25 e 50 metros, representando 95% e 99% dos valores, respectivamente. No caso de Dublin, os limites são 300 e 350 metros, representando 90% e 95% dos valores.

## Tempo por trajetória



As CDFs de tempo por trajetória nos mostram que no caso de Dublin há uma curva um pouco mais acentuada, indicando uma porcentagem maior de valores pequenos. Em Roma, pouco mais de 70% dos valores estão abaixo de 50 minutos, já em Dublin isso cobre entre 85% e 90% dos valores.

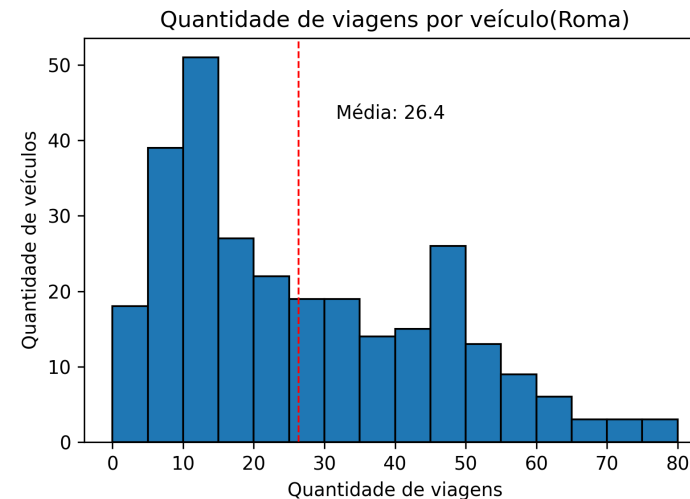
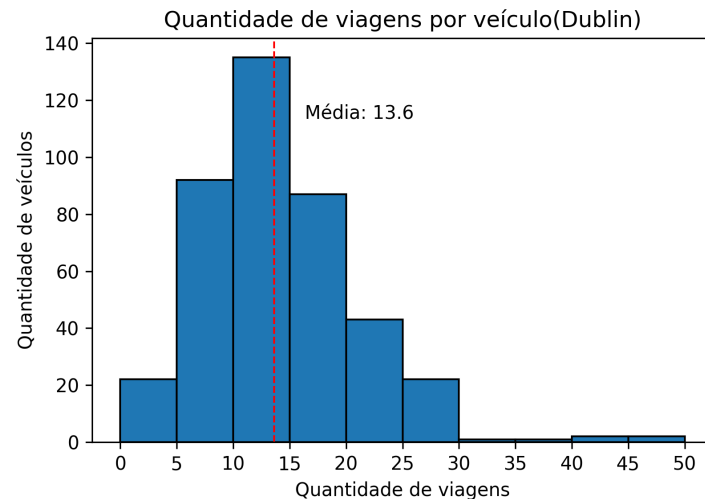
## Trajетórias por veículo



As CDFs de trajetórias por veículo mostram grande diferença de comportamento entre as duas bases de dados. No caso de Roma, praticamente 100% dos veículos possuem menos de 30 trajetórias em um único dia. Dublin mostra valores muito maiores, indo até 200 trajetórias por veículo num único dia, onde 30 trajetórias representa menos de 50% do total. O provável motivo é que os táxis ficam muito tempo parados esperando por passageiros, já os ônibus possuem horário e rotas fixadas, o que explicaria a maior quantidade de trajetórias por veículo. Além disso, cada ponto intermediário de um ônibus pode ser considerado como fim de uma trip e início de uma nova pelo algoritmo de detecção de viagens.

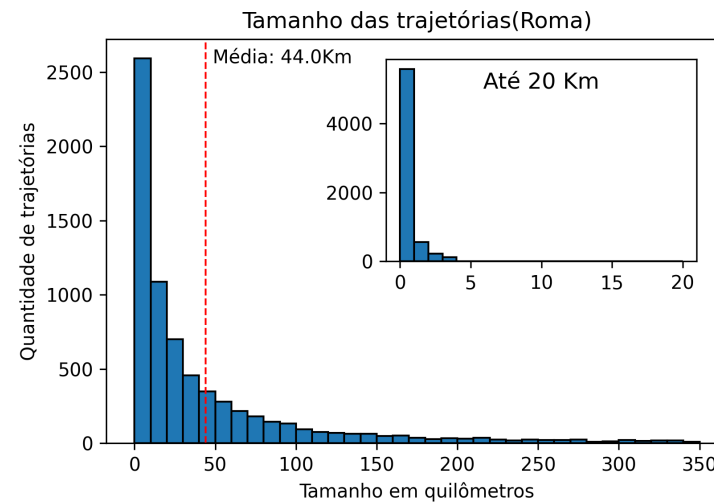
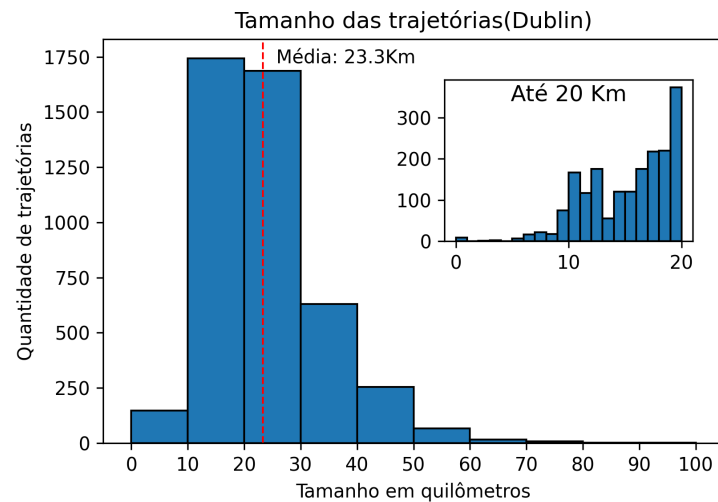
# Histogramas

## Viagens por veículo



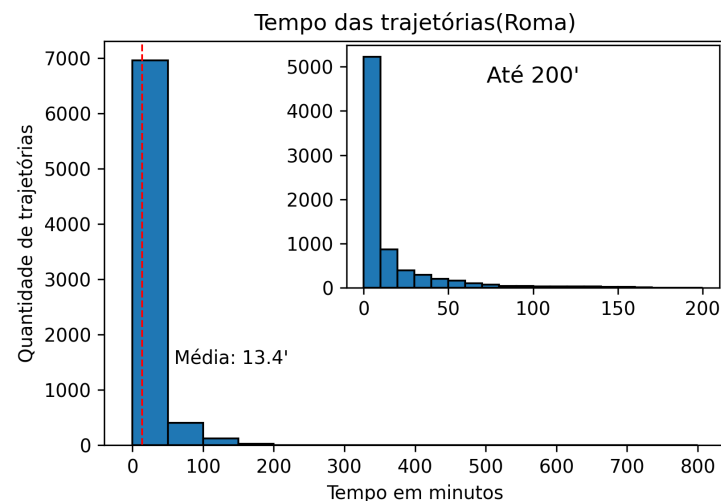
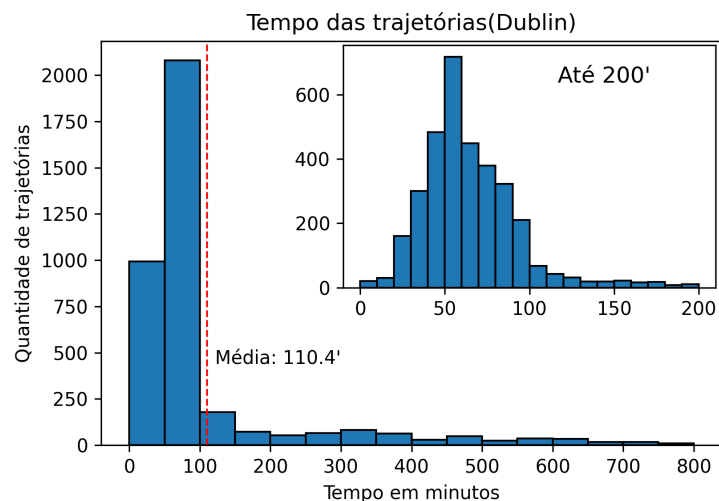
Os gráficos de viagens por veículo mostram quantos veículos possuem determinada quantidade de viagens. Nesse caso, as viagens dos ônibus são consideradas apenas pontos iniciais aos finais, desconsiderando paradas. Dessa forma vemos uma quantidade média menor de viagens para os veículos de Dublin, com uma concentração grande de veículos com até 30 viagens por dia. Roma mostra uma distribuição mais espaçada, contendo veículos com até 80 viagens e uma média de 26 viagens por veículo, praticamente o dobro da média de Dublin. Como ônibus são veículos com trajetórias normalmente longas, isso explicaria a quantidade menor de viagens diárias, afinal, os táxis vão do ponto A ao B na trajetória mais rápida possível, o que pode não ser verdade para muitas trajetórias de ônibus.

## Tamanho das trajetórias



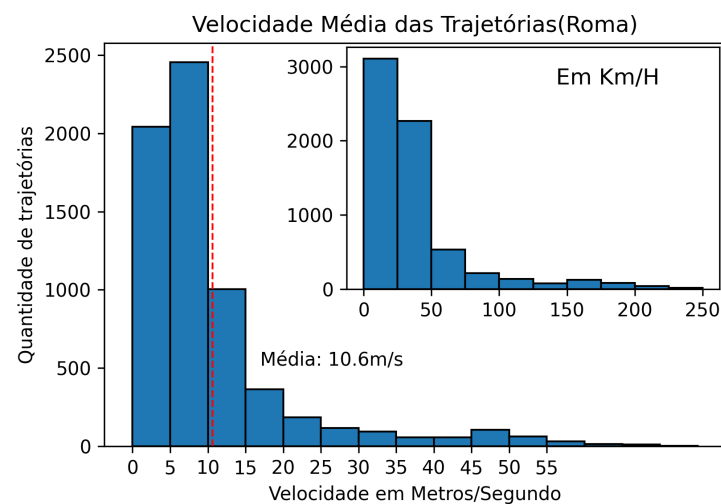
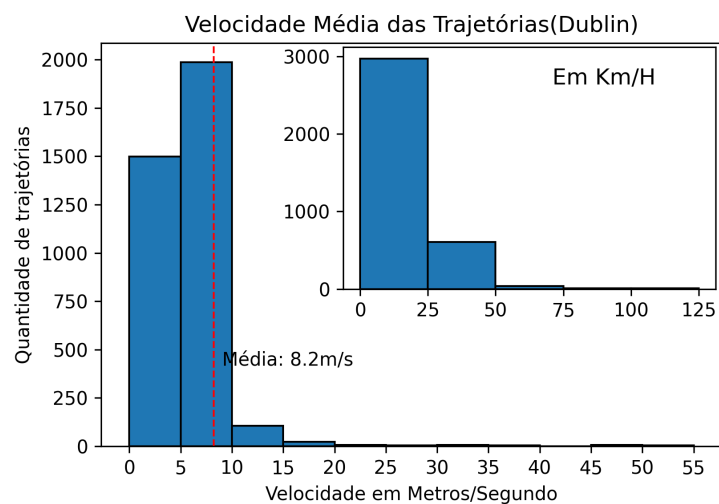
Estes histogramas informam a quantidade de trajetórias que possuem determinado tamanho. Roma possui um comportamento de cauda longa, com grande quantidade de trajetórias pequenas, a maioria menor que 5Km, porém com uma média de 44Km. Já Dublin se aproxima mais de uma distribuição F com média de 23Km e poucas trajetórias abaixo de 5Km. Isso pode ser explicado pela diferença comportamental dos veículos. Ônibus possuem trajetórias já traçadas e planejadas, sendo incomum viagens pequenas pois seriam ineficientes. Táxis são comumente usados por pessoas que precisam de deslocamento rápido, o que seria mais apto a ser escolhido para trajetórias pequenas ou muito grandes.

## Tempo das trajetórias



Os histogramas de tempo das trajetórias representam quantas trajetórias estão incluídas em determinado intervalo de tempo. Roma mostra possuir trajetórias muito curtas com média de 13 minutos, e, novamente, é observável o comportamento de cauda longa quando estabelecemos o limite do gráfico de até 200 minutos. Dublin possui uma média de tempo muito maior, 110 minutos, e mantém o comportamento de distribuição F. É um padrão de comportamento esperado para ônibus que tendem a ter trajetórias longas.

## Velocidade média das trajetórias



Os histogramas de velocidade média das trajetórias expõem quantas trajetórias possuem uma velocidade média dentro de um intervalo específico. A diferença entre as médias não é muito grande entre as duas bases, sendo a de Roma pouco maior. Sendo ônibus veículos de rotação interna na cidade, é esperável que a velocidade média seja mais baixa, além de controlada, o que pode ser confirmado pelo gráfico, com a maior parte das trajetórias de velocidade média até 50Km/H. Os táxis por terem mais liberdade de movimentação, tendem a se mover mais rápido, o que também é demonstrado nos gráficos, com uma quantidade um pouco maior de trajetórias cuja média de velocidade é acima dos 50Km/H