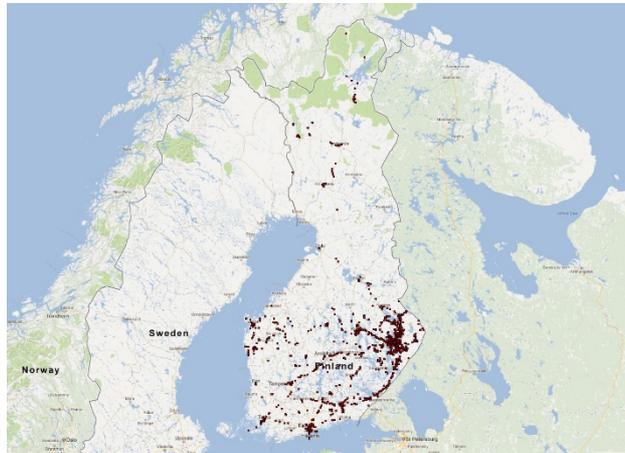


Exercício 5: OPTIONAL

Não deve ser entregue relatório.

- 1. Análise de utilizadores de um jogo online na Finlândia:** a seguinte imagem mostra a localização de utilizadores de um jogo online na Finlândia. Pretende-se agrupar os jogadores para se encontrar a melhor localização para se instalar novos servidores.



- Utilizando o algoritmo de clustering k-means nos dados em anexo, agrupe os utilizadores em 2 e em 3 grupos geográficos para 10 iterações e centroides iniciais aleatórios.
 - Apresente os resultados em dois gráficos, na mesma figura, para $k=2$ e $k=3$, em que cada grupo no gráfico tem a sua cor. Faça os gráficos anteriores em cima do mapa da Finlândia.
 - Faça o gráfico da soma do quadrado das distâncias aos centroides em função do k . Discuta o melhor valor de k .
- 2. Redução de dimensionalidade em características de semicondutores.**
- Considerando os dados no ficheiro em anexo (semiconductors.csv), aplique o método de Principal Component Analysis (PCA) para encontrar as duas primeiras componentes principais. Faça uma tabela com os vetores próprios das duas componentes principais.
 - Faça o gráfico da componente principal 2 (PC2) em função da componente principal 1 (PC1).
 - Aplique o k-means e classifique os semicondutores em diferentes grupos.
- 3. Mapeamento do cérebro de um rato:** Considere o seguinte dataset do mapeamento de ligações no cérebro de um rato com +21000 ligações (<http://networkrepository.com/bn-mouse-brain-1.php>). Cada linha tem um nodo de origem e um de chegada da ligação.
- Calcule o maior e menor grau de ligação.
 - Calcule o grau de ligação médio.
 - Calcule o cluster coefficient.
 - Calcule a média dos caminhos mais curtos a partir do primeiro nodo (nodo 0).