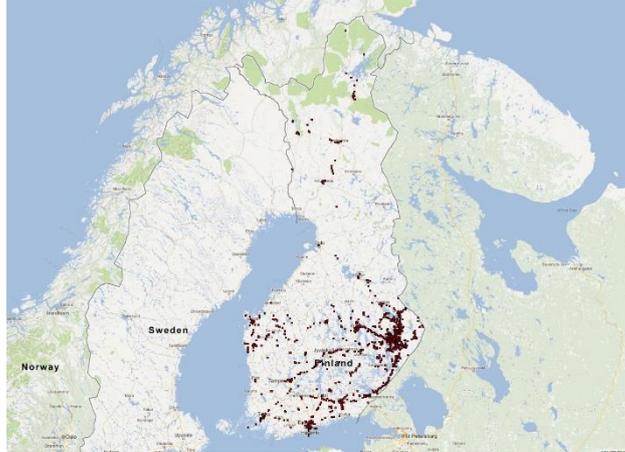


Exercício 10: Análise de dados (parte III)

Deve ser entregue relatório até a próxima aula.

- 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.
 - (Opcional)** Para $k=4$ apresente 4 gráficos na mesma figura para 2, 3, 4, e 5 iterações. Discuta o efeito das iterações.
 - Faça o gráfico da soma do quadrado das distâncias aos centroides em função do k . Discuta o melhor valor de k .
 - (Opcional)** Faça os gráficos anteriores em cima do mapa da Finlândia.
- 2. (Opcional) Análise de imagens.** Usando as imagens em anexo:
- Use o k-means para espaço de cores e fazer clustering de cores.
 - Faça o mesmo no espaço LAB.
 - Use o k-means em escala de cinzas e faça clustering de pixels.
- 3. 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).
 - (Opcional)** Aplique o k-means e classifique os semicondutores em diferentes grupos.