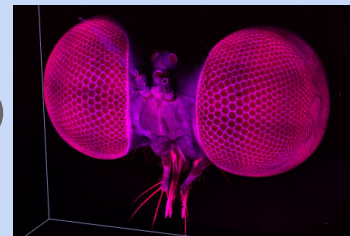# Valentina Staneva - Data Scientist, eScience Institute, University of Washington

PhD - Object Segmentation and Tracking
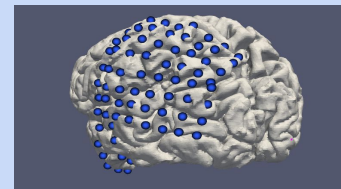
three.js

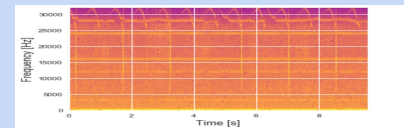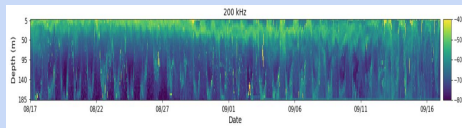Pipelines for Automatic Cornea Image Segmentation

plotly

C++

Spatiotemporal Data
Optimization
Dimensionality
Reduction

R

python

aws

docker

SQL

Data Science for Social Good: transportation projects

Multimodal Neural Encoding: ecog + video + audio

Ocean Acoustic Data Mining: echosounder + hydrophone

Software Simplicity vs Scalability

Need to handle large datasets: images, streaming time series, videos, networks, ...

Need to have libraries:
➢ Easy to use
➢ Easy to develop for
➢ Work the same on small and large datasets
➢ Work the same on laptop, university cluster, cloud, gpu

Some examples: dask, keras, jupyterhub, kubernetes, ...

Should we rewrite the libraries, should we change the deployment systems?