Here is a brief description for three code samples that I submitted. Please find more details in the readme.md under each repository.

**Sample 1 Visualizations**

* **File:** Sample1\_Visualization.zip
* **Code:** <https://github.com/DeyunYinWIPO/visualizations>
* **Description:** This sample is a collection of code for visualizing global innovation activities with maps and streamgraphs. It is produced for WIPO's World Intellectual Property Report 2019 & 2021
* **Report:** WIPO (2019). *World Intellectual Property Report 2019: The Geography of Innovation: Local Hotspots, Global Networks*. World Intellectual Property Organization.
* **Note:** Please note as the size of generated charts are very large, I didn’t enclose them into Sample1\_Visualization.zip. Please kindly check the github link for the results.

**Sample 2 Machine Learning**

* **File:** Sample 2\_Machine learning.zip
* **Code:** <https://github.com/DeyunYinWIPO/Machine-learning>
* **Description:** This is a sample code for predicating innovation performances based on machine-learning algorithms

**Sample 3 Novelty Indicators**

* **File:** Sample 3\_Novelty.zip
* **Code:** <https://github.com/DeyunYinWIPO/Novelty>
* **Description:** This code is for Sotaro Shibayama, Deyun Yin, and Kuniko Matsumoto (2021). Measuring Novelty in Science with Word Embeddings. PLOS ONE.(forthcoming)
* **Abstract:** Novelty is a core value in science, and a reliable measurement of novelty is crucial. This study proposes a new approach of measuring the novelty of scientific articles based on both citation data and text data. The proposed approach considers an article to be novel if it cites a combination of semantically distant references. To this end, we first assign a word embedding – a vector representation of each vocabulary – to each cited reference on the basis of text information included in the reference. With these vectors, a distance between every pair of references is computed. Finally, the novelty of a focal document is evaluated by summarizing the distances between all references. The approach draws on limited text information (the titles of references) and publicly shared library for word embeddings, which minimizes the requirement of data access and computational cost. We share the code, with which one can compute the novelty score of a document of interest only by having the focal document's reference list. We validate the proposed measure through three exercises. First, we confirm that word embeddings can be used to quantify semantic distances between documents by comparing with an established bibliometric distance measure. Second, we confirm the criterion-related validity of the proposed novelty measure with self-reported novelty scores collected from a questionnaire survey. Finally, as novelty is known to be correlated with future citation impact, we confirm that the proposed measure can predict future citation.