**Lab 4 Report**

**Title:** Compare the Interpolation Methods

**Notice:** Dr. Bryan Runck

**Author:** Liang-Ting Chen

**Date:** Apr 19, 2021

**Project Repository:**

<https://github.com/chen6761/GIS5572>

**Abstract**

In the lab, I built an ETL to download the current 30 days data from the NDAWN site and mapped them. To show the interpolation, I followed the instruction to analyze the data with three different methods based on the ESRI decision guide and do the comparison and contrast.

**Problem Statement**

Interpolate the weather data in the current 30 days

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **Requirement** | **Defined As** | **Spatial Data** | **Attribute Data** | **Dataset** | **Preparation** |
| 1 | Weather data | Raw input from NDAWN | Local Temperature |  | [NDAWN](https://ndawn.ndsu.nodak.edu/) |  |

**Input Data**

In the lab, I need weather data to interpolate and make maps.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | Weather data | As the input to be interpolate | [NDAWN](https://ndawn.ndsu.nodak.edu/) |

**Methods**

**ETL**

I referred to my code in lab 1 to download the data first. To set the variable for getting the current data, I used the datatime module as Jeffrey did. But the difference is that I just set it in my data URL and it could work when parsing the website. (Figure 1)

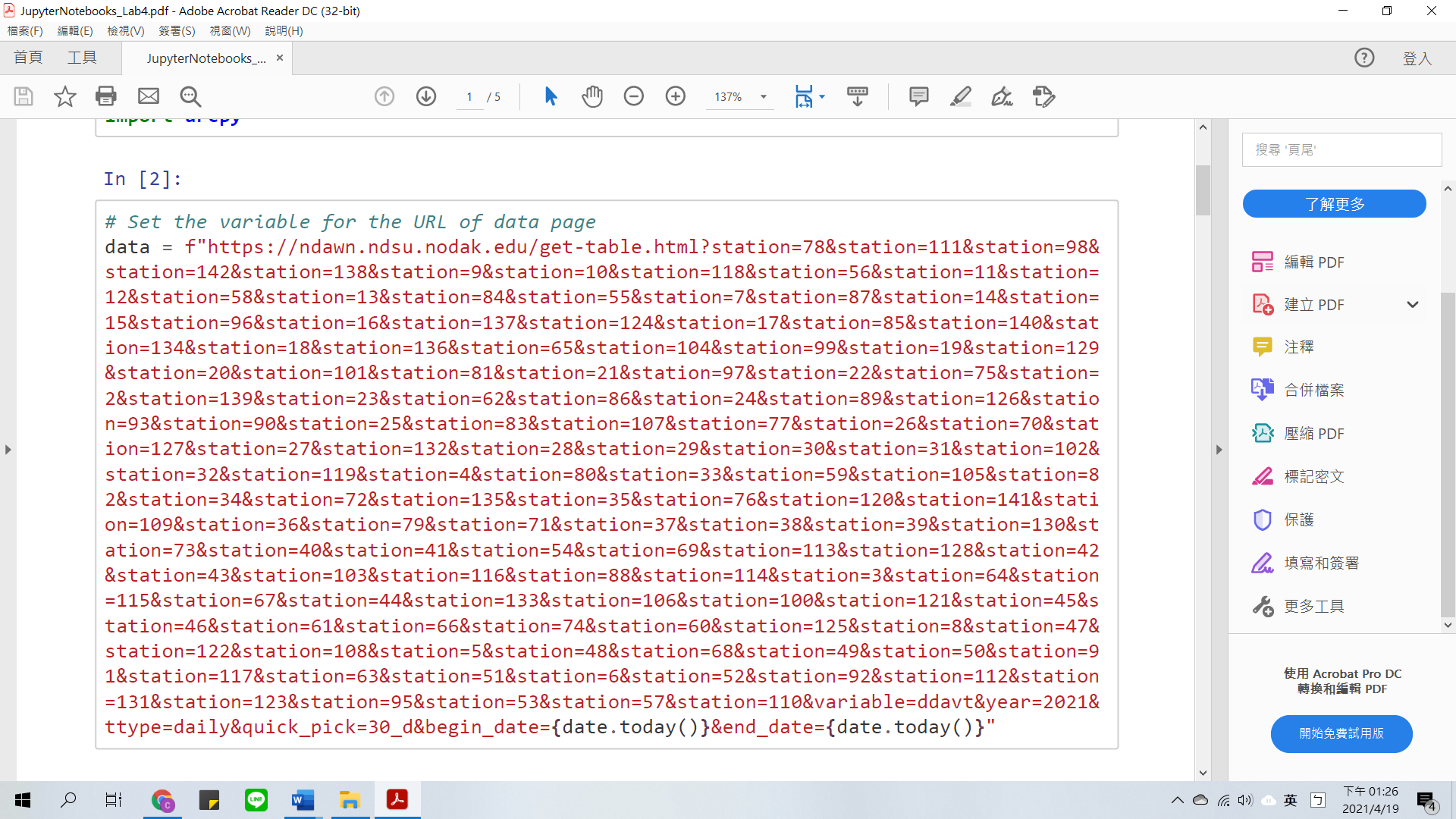


Figure 1

After downloading the data, I found that there are several unnecessary rows in the CSV file. So, I turned it into the data frame and cleaned it up by leaving the useful parts. I read the CSV since the 4th row to avoid the background information for the first three rows. Then, I renewed the column, separated the content with commas, renewed the header, and dropped some useless columns. (Figure 2)

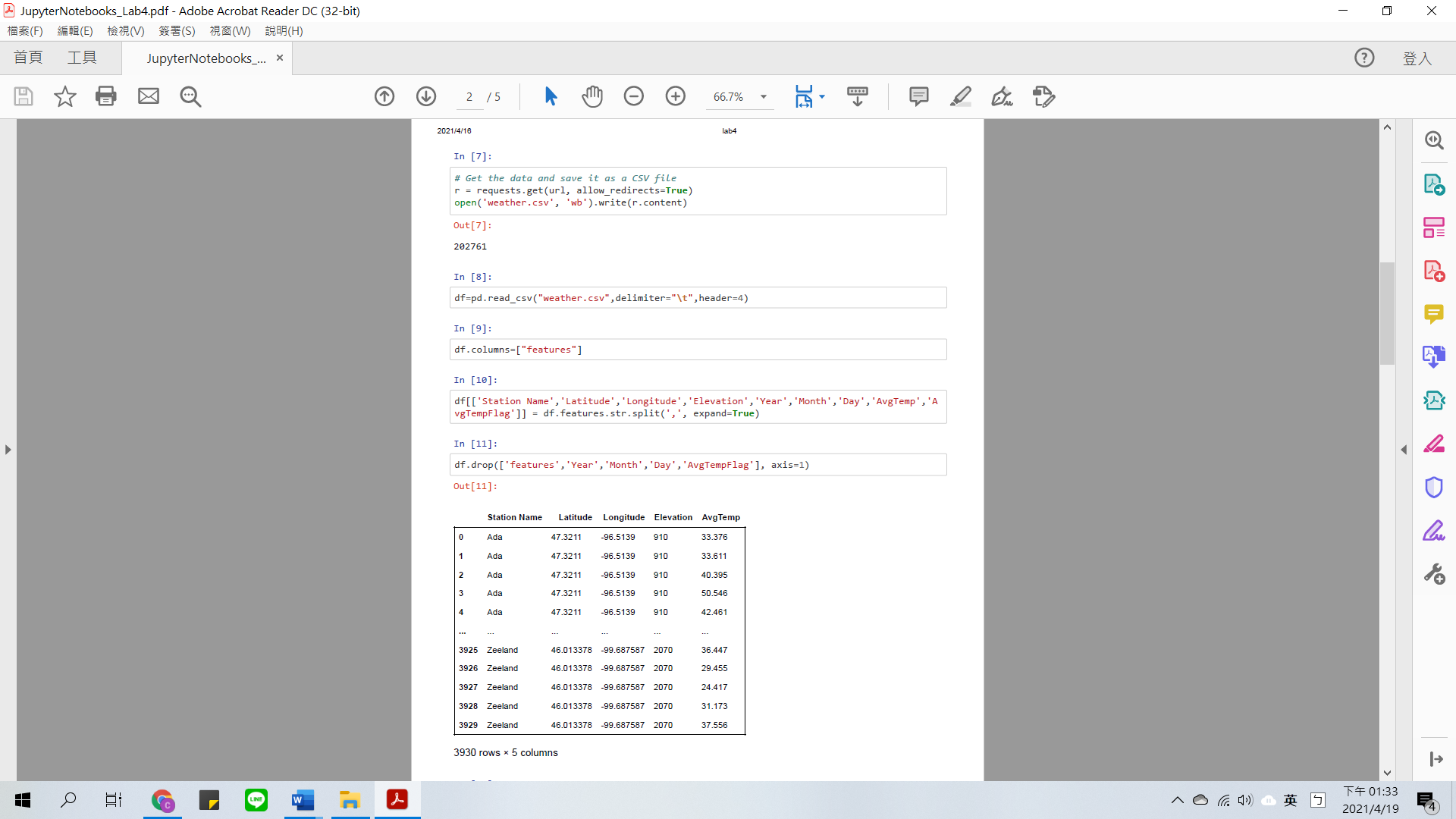


Figure 2

The last step was to change the data type of temperature to make it calculable. I grouped them up, calculated the average temperature, and saved it as a new CSV file for locating the points in ArcGIS Pro. (Figure 3)

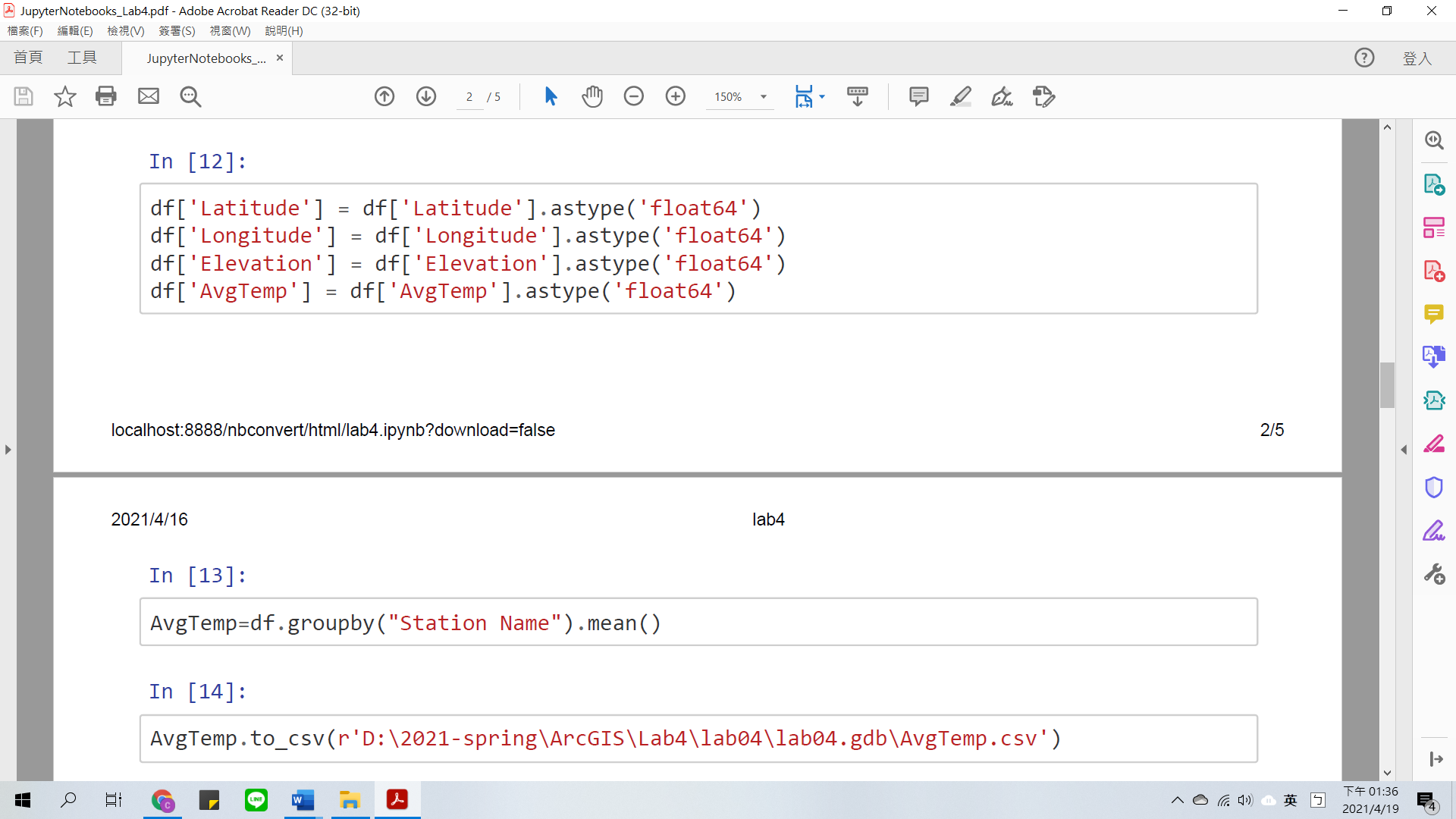
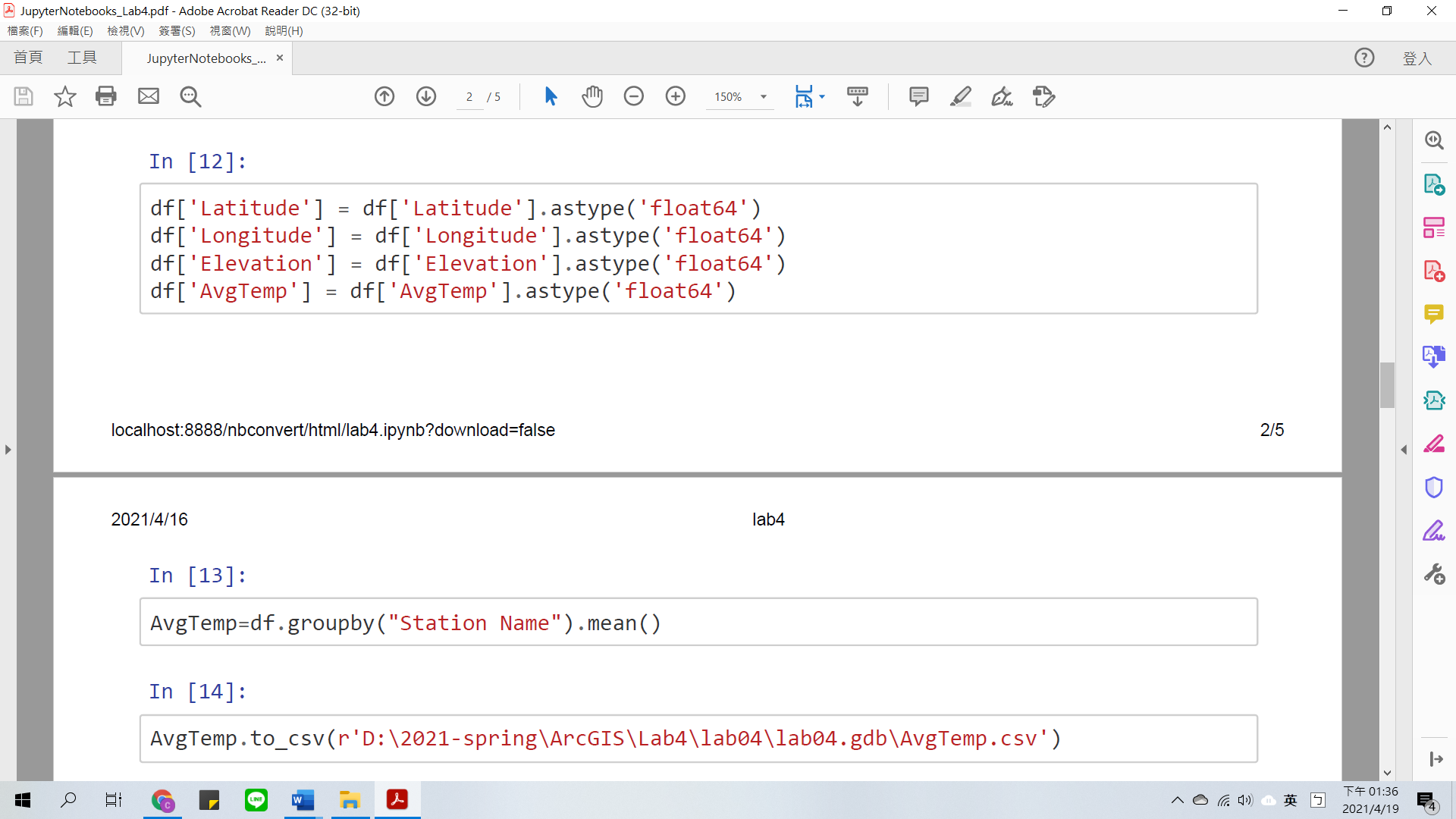


Figure 3

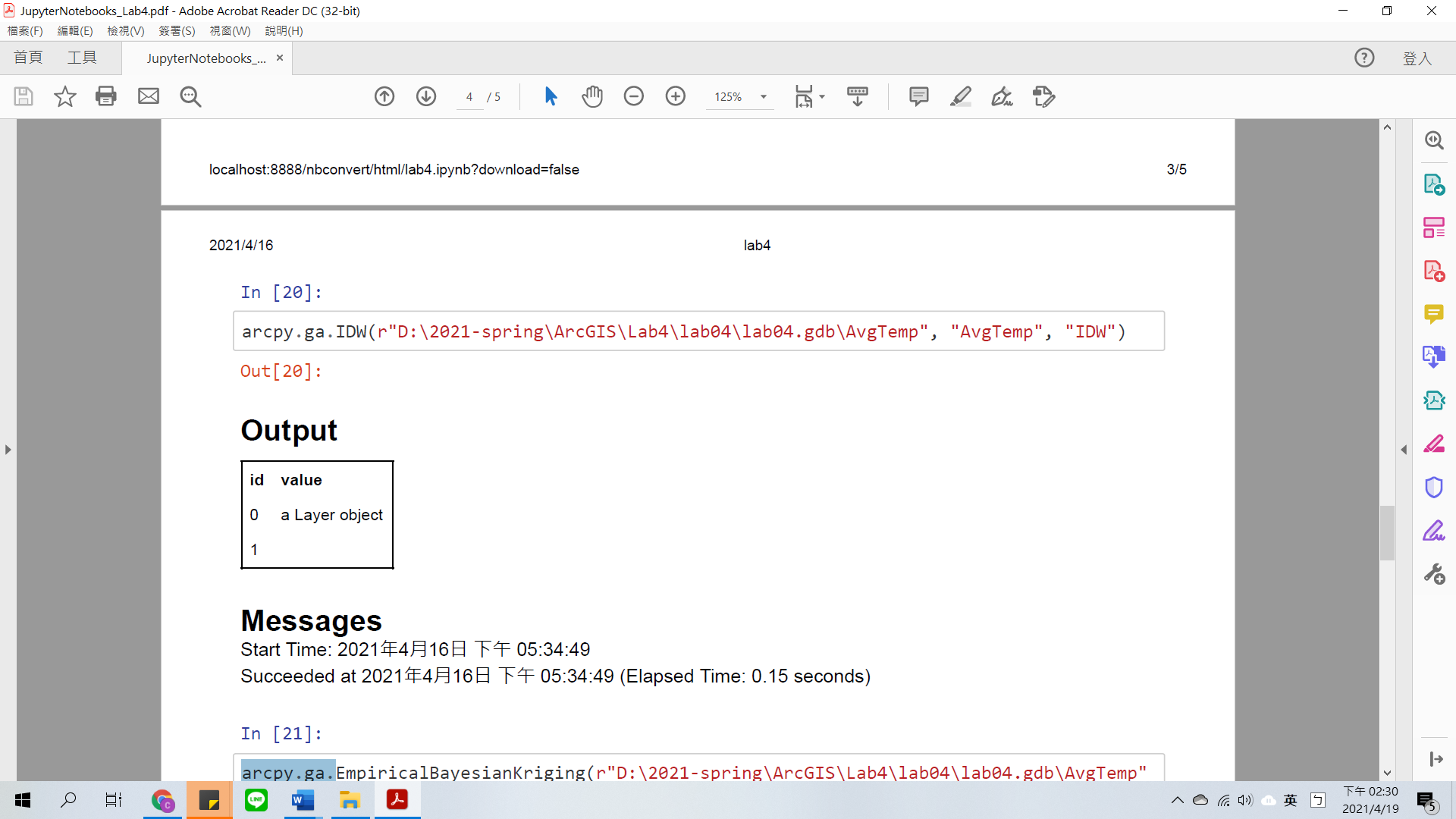
**Interpolation**

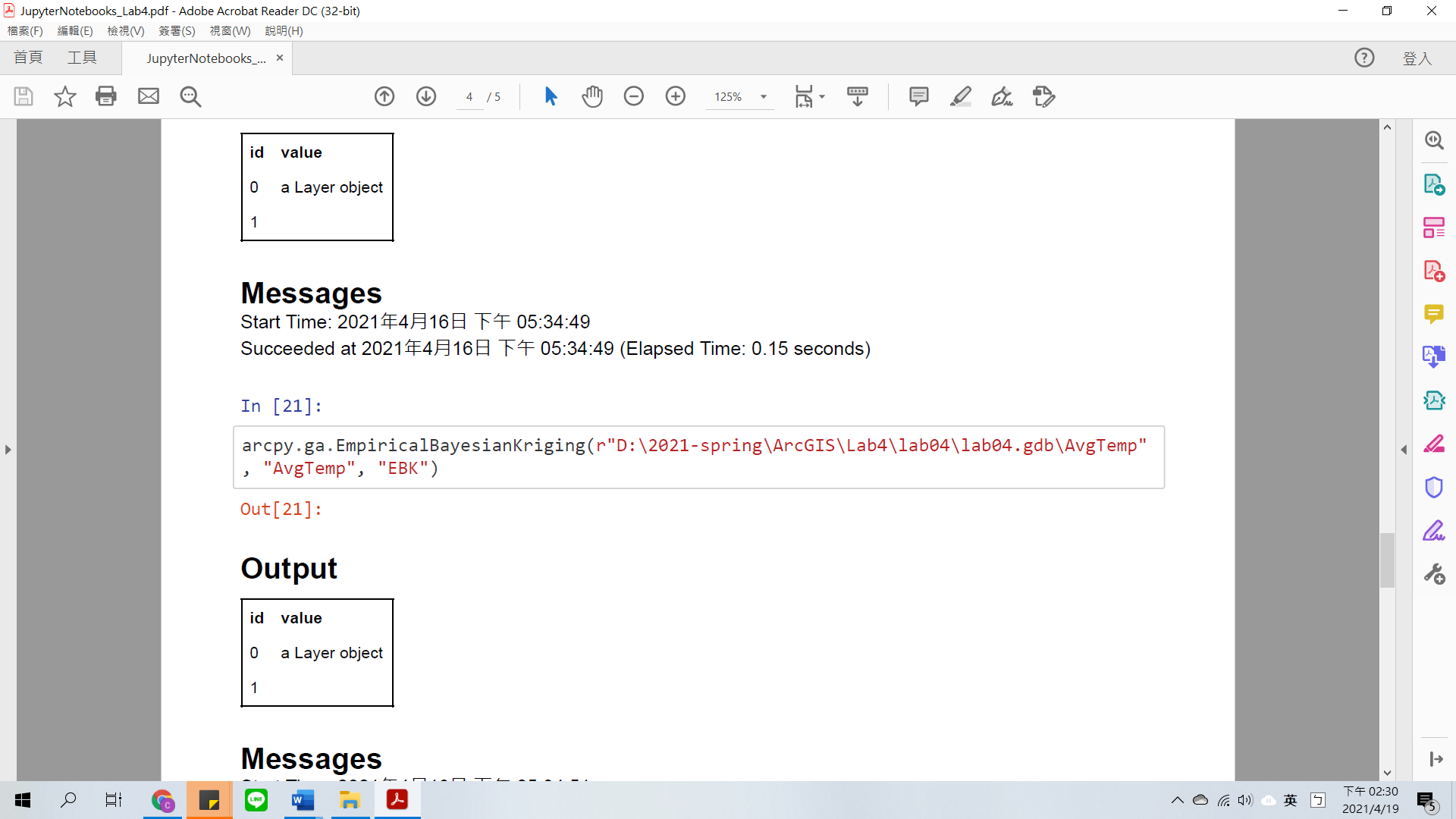
Follow the ESRI decision guide and the instruction, I selected the Inverse Distance Weighted (IDW), Empirical Bayesian Kriging (EBK), and Radial Basis Function (RBF) as my interpolation methods. I used IDW according to the lab requirement, and the reason that I chose the other two methods is below.

Empirical Bayesian Kriging (EBK) can estimate the parameter value automatically, besides, it is also one of the most suitable interpolation methods when the focused observed distribution is unsure. (Dhakal et al., 2020)

While Radial Basis Function (RBF) is a way to interpolate does not consider the external variable, but in some article shows that it exist lower Root mean square error (RMSW) in interpolating the temperature data. (Piazza et al., 2015)

To analyze the temperature data with these interpolation methods, I used arcpy.ga syntax in the Jupyter Notebook. (Figure 4) Then, I used .listLayouts and .exportPDF to get the maps.





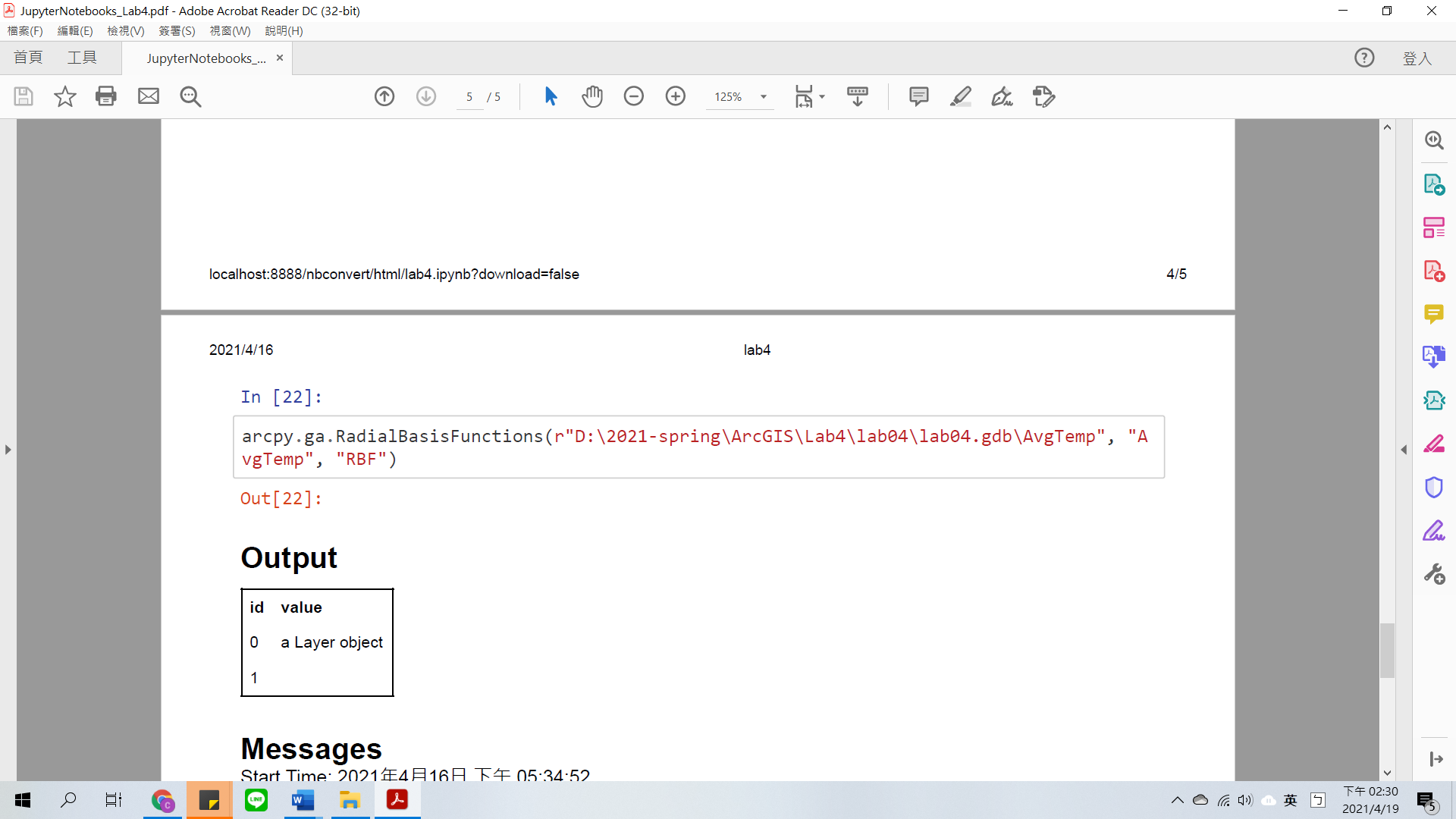
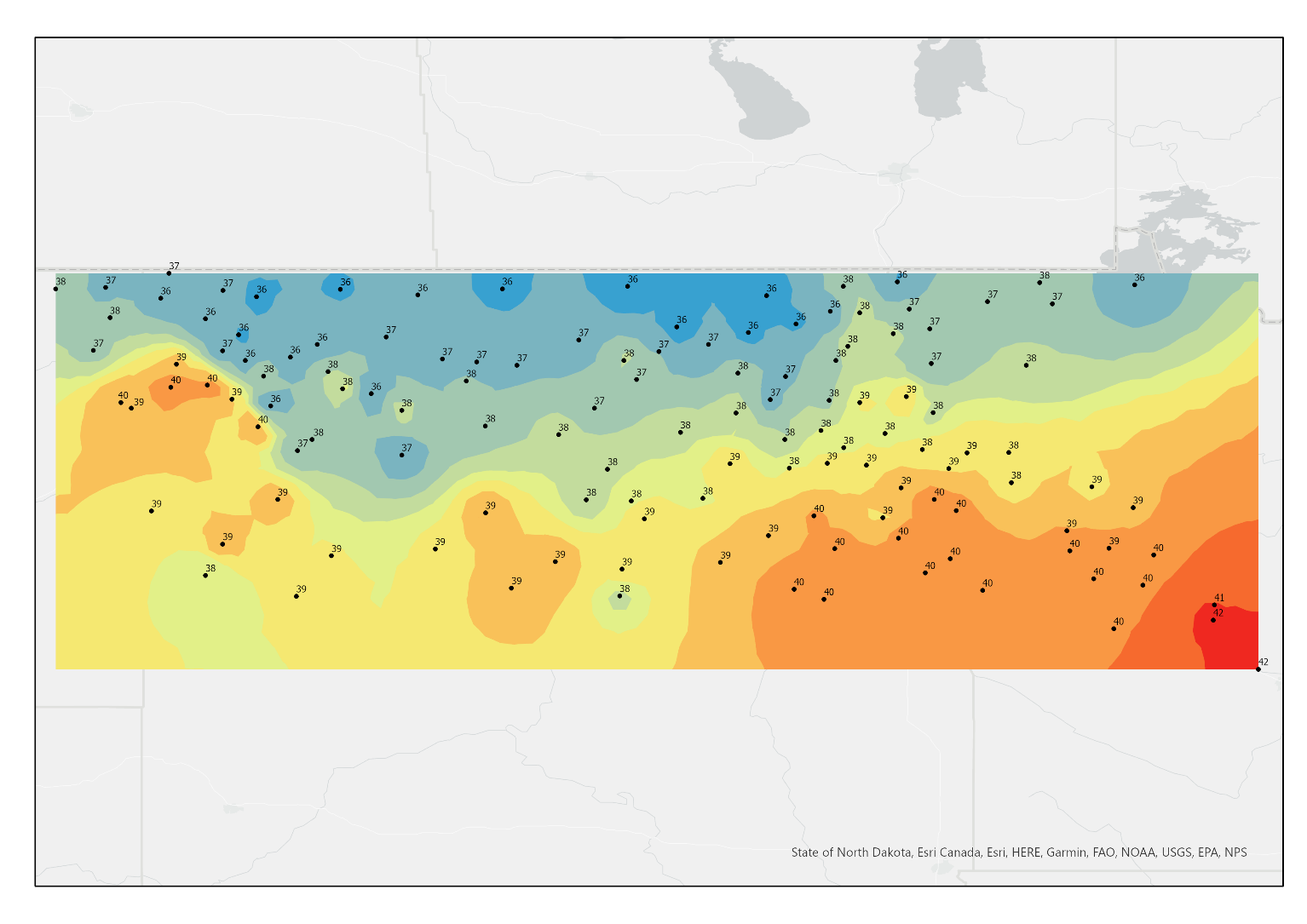


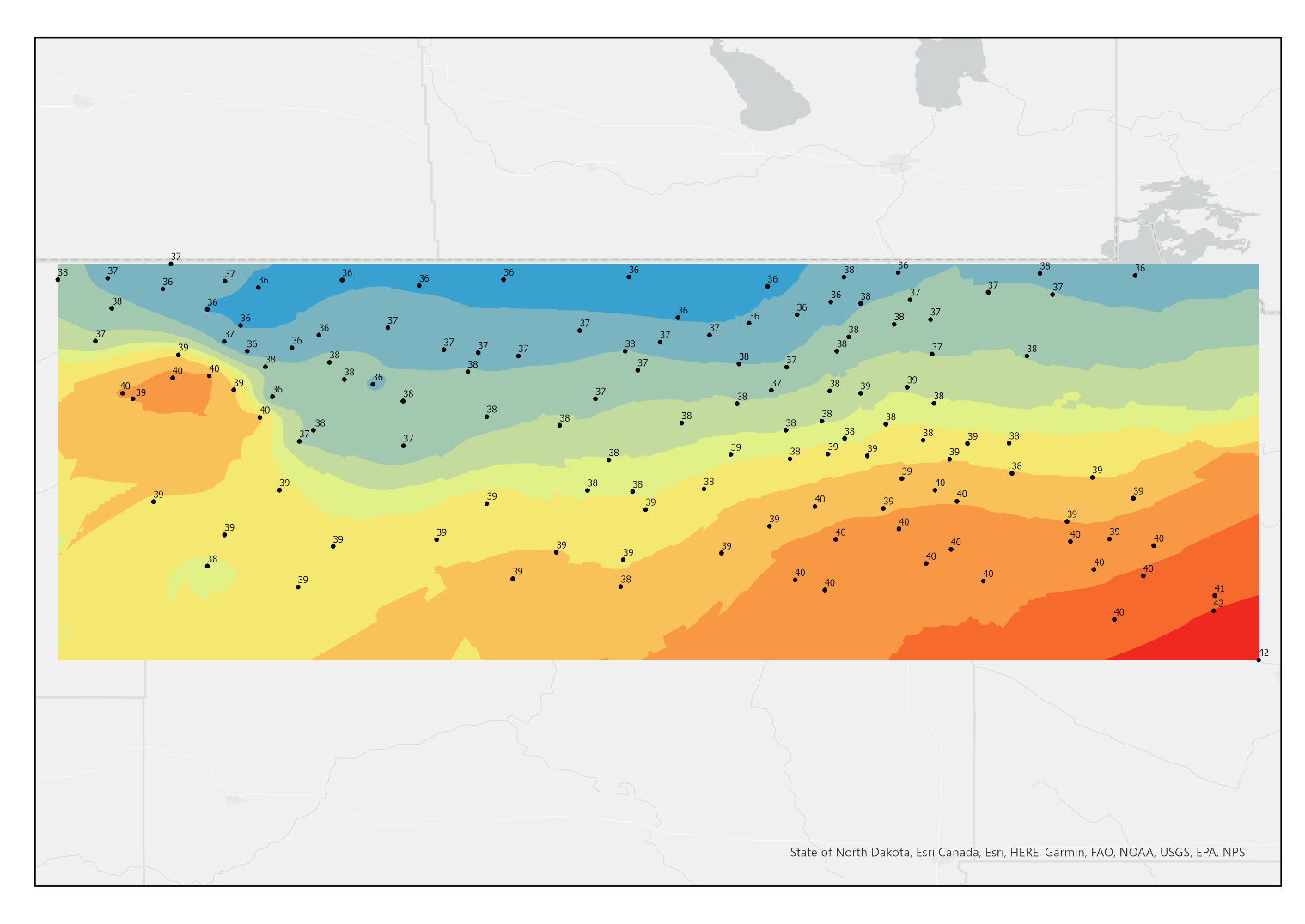
Figure 4

**Results**

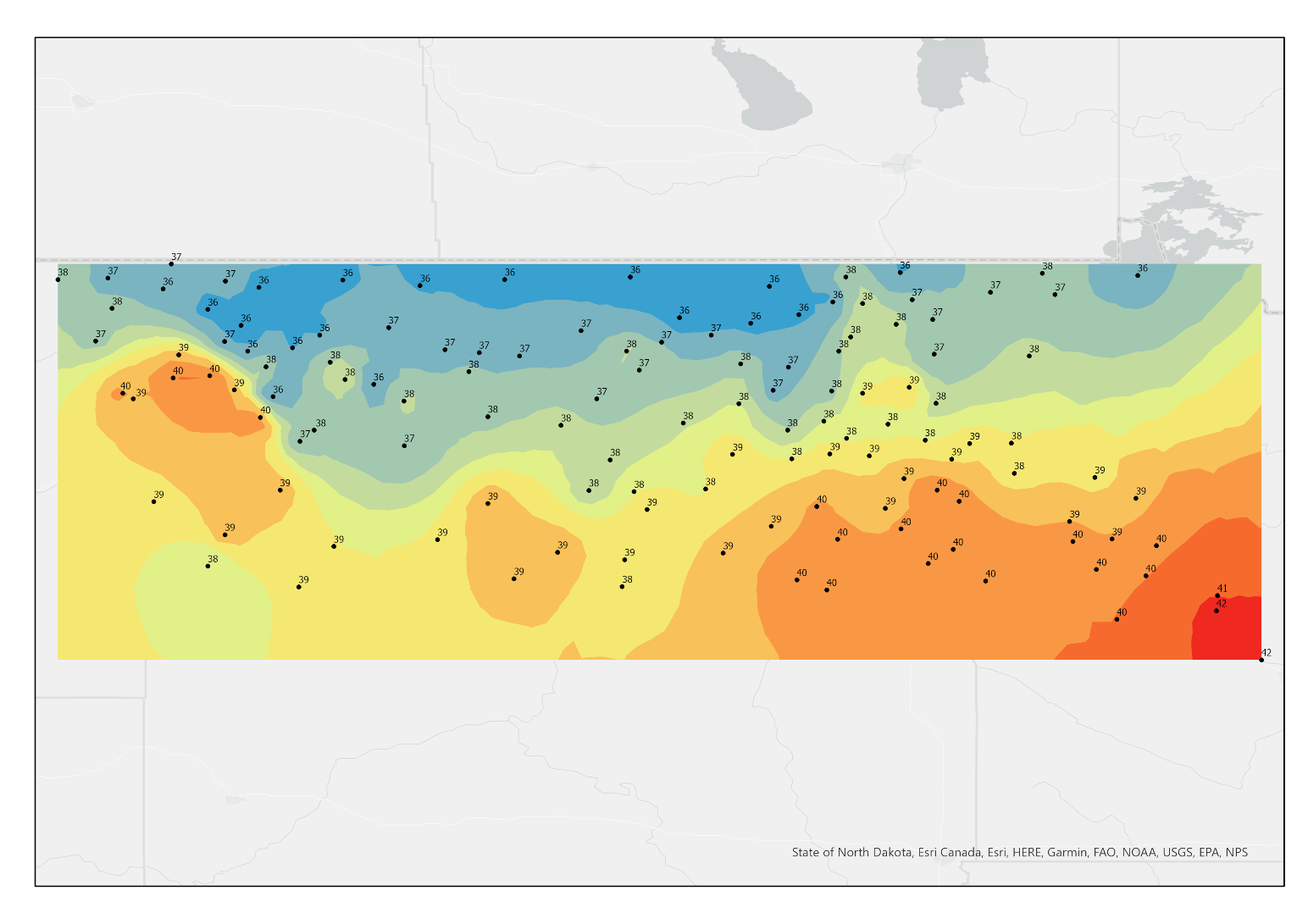
**IDW**

****

**EBK**

****

**RBF**

****

**Results Verification**

I found most of the local temperatures were between 36 to 42, and all of the interpolation maps showed the graduated color based on the temperature. None of the temperatures was assigned to the wrong color, so I thought my results were correct.

The IDW and the RBF looked similar, but I thought the IDW was better. It showed the best interpolation because the graduated color was presented suitably based on the distribution of temperature. It considered more detailed data and made changes nearby them. The RBF was also good, but it ignored some of the sparse data and did not change the graduated color. The EBK only showed the color based on the latitude. Although it presented most of the data in a suitable color, it did not make the color change appropriately between stations.

**Discussion and Conclusion**

The lab makes me understand more about interpolation functions in ArcGIS. I also found that the different interpolation methods are suitable for different kinds of data. It is important to understand my data property first to choose the appropriate one.

**References**

*Dhakal, K., Kakani, V. G., Ochsner, T. E., & Sharma, S. (2020). Constructing retrospective gridded daily weather data for agro‐hydrological applications in Oklahoma. Agrosystems, Geosciences & Environment, 3(1), 6. https://doi.org/10.1002/agg2.20072*

*Piazza, A., Conti, F., Viola, F., Eccel, E., & Noto, L. (2015). Comparative Analysis of Spatial Interpolation Methods in the Mediterranean Area: Application to Temperature in Sicily. Water, 7(12), 1866–1888. https://doi.org/10.3390/w7051866*

**Self-score**

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Description** | **Points Possible** | **Score** |
| **Structural Elements** | All elements of a lab report are included **(2 points each)**:  Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score | 28 | 28 |
| **Clarity of Content** | Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**. | 24 | 23 |
| **Reproducibility** | Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified. | 28 | 27 |
| **Verification** | Results are correct in that they have been verified in comparison to some standards. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**. | 20 | 19 |
|  |  | 100 | 97 |