

# **DS4D Group 14**

**Encyclopedia Britannica**

**Dake Wang, Keren He, Qiuyue Ding, Shijie Chen, Yuhan Ma**

# **Encyclopedia Britannica**

# #1 Using Topic Modeling and LDA to analyze Encyclopedia

Topics in LDA model



## #2 Name each topic

### Topic 8: Trade

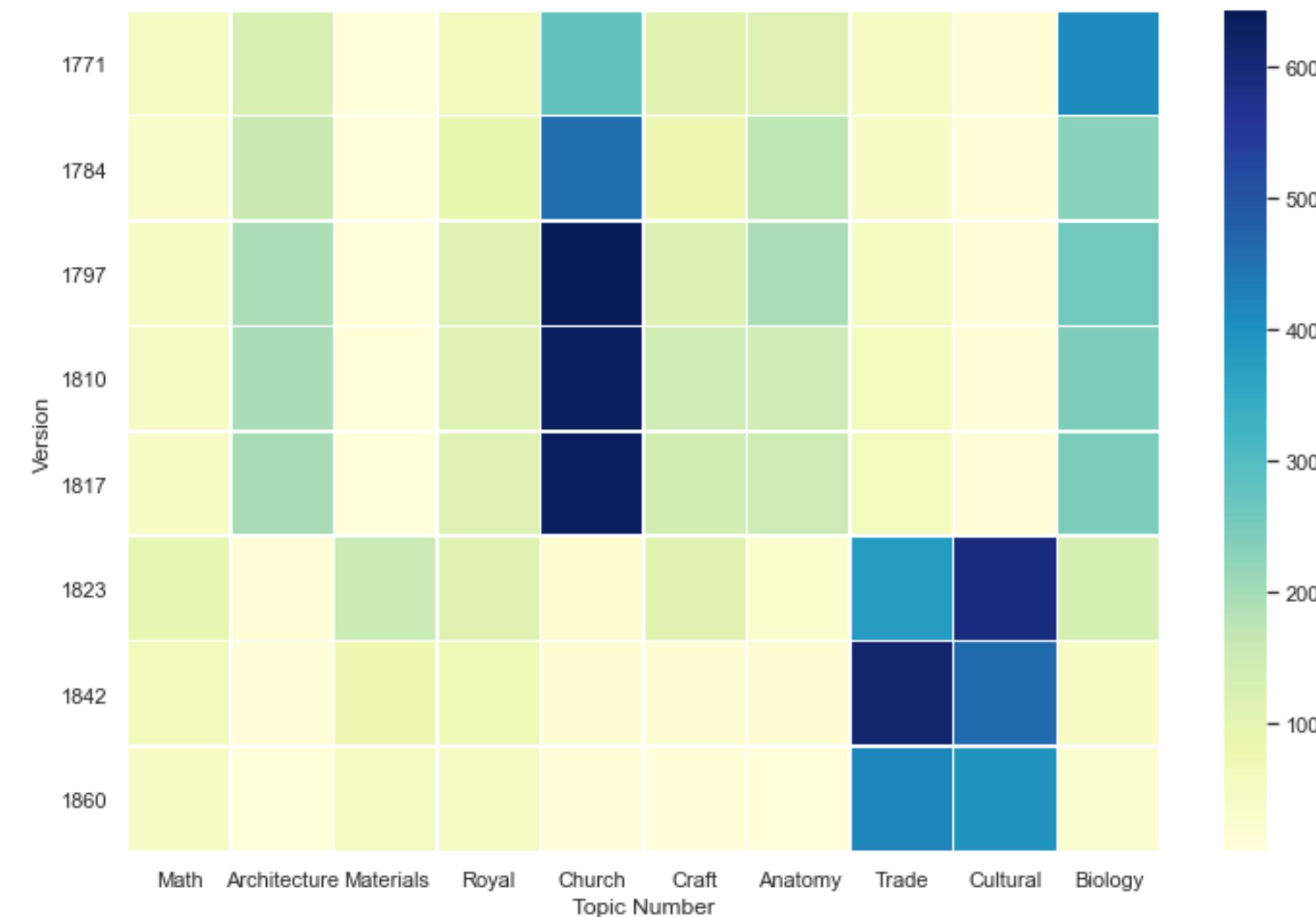
south, feet, east, **trade**, county, west, land, **island**, small, population, considerable, built, coast, state, situated, mountains, houses, **value**, contains, government, islands, british, extent, london, nearly, rivers, **bank**, **market**, **money**, house, chief, chiefly, building, produce, **corn**, soil, present, various, district, numerous, square, english, towns, extensive, western, northern, states, course, eastern, **cotton**

sample#1: a town situated on the isthmus of Panama. See Panama.PORTO FERRAJO, the capital of the Island of Elba, in the province of Pisa, and the duchy of Tuscany<sup>^</sup> celebrated as the residence of Napoleon during his banishment to that island. It is situated on a tongue of land running into the sea, and forming a small bay. It is fortified, and contains two churches, an hospital, 600 houses, with 3120 inhabitants, who depend chiefly upon some salt works and the tunny fishery. Lat. 42. 49. 6. Long. 9. 20. 1 ...

sample#2: or Newburgh, a borough-town of North Wales, in the island of Anglesey, and hundred of Menai, 257 miles from London, and twelve from Beaumaris. It was once the residence of the princes of Anglesey, and a corporation founded by Edward I. There is a market, which is held on Tuesday. The population amounted in 1801 to 599, in 1811 to 750, in 1821 to 756, and in 1831 to 804.NEW BRUNSWICK, a British province of North America, situated between the parallels of 45. 5. and 48. 4.30. north latitude, and ...

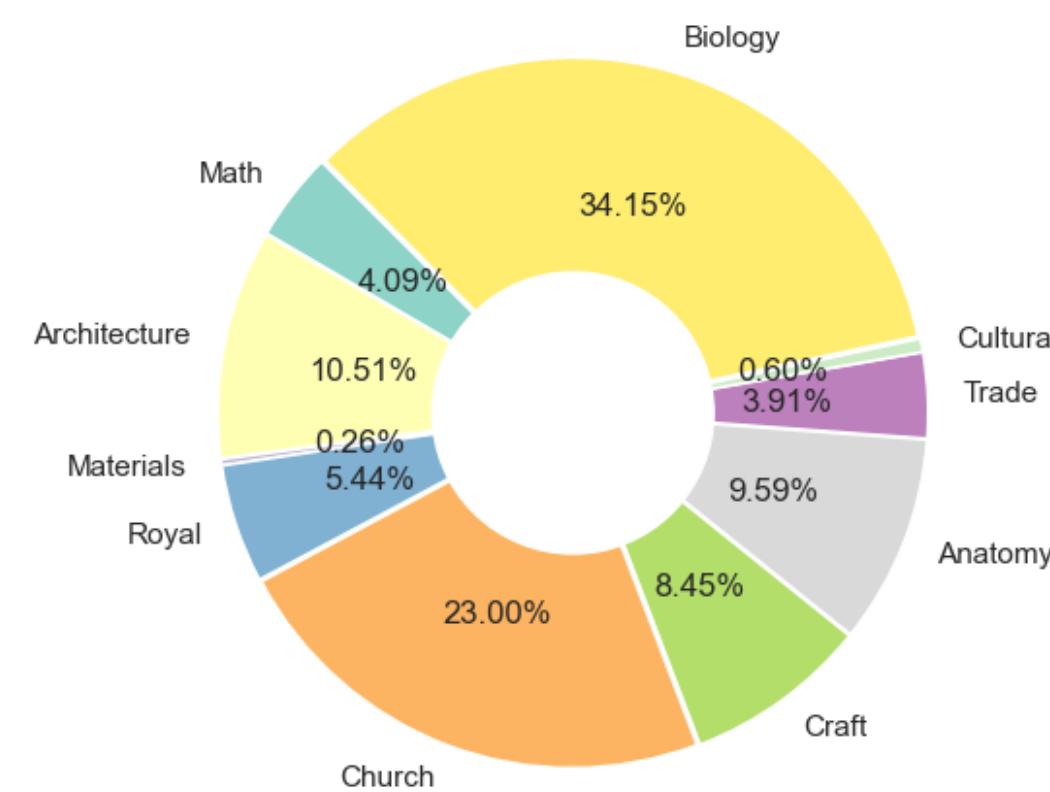
sample#3: or Carmarthenshire (Welsh Caerfyrddiri), a maritime county in South Wales, is bounded on the north by Cardigan, on the east by Brecon, on the south by Glamorgan and the Bristol Channel, and on the west by Pembroke. Its greatest length is from S.W. to N.E., about 52 miles; its greatest breadth, S.E. to N.W., about 28 miles. It possesses an area of 947 square miles, or 606,331 acres, and is thus the largest of all the Welsh counties. It contains 77 parishes, and is in the diocese of St David's. I ...

# #3 Analyze the relationship between each edition and topic

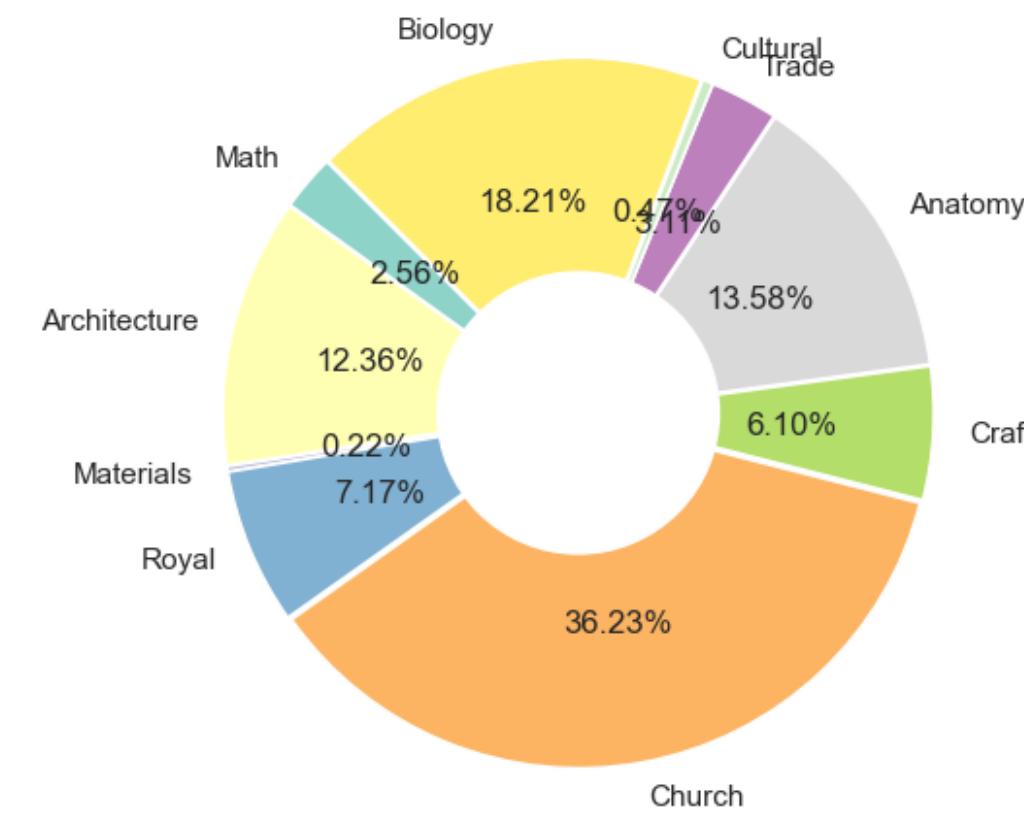


Frequency Distribution of Each Topic by Edition

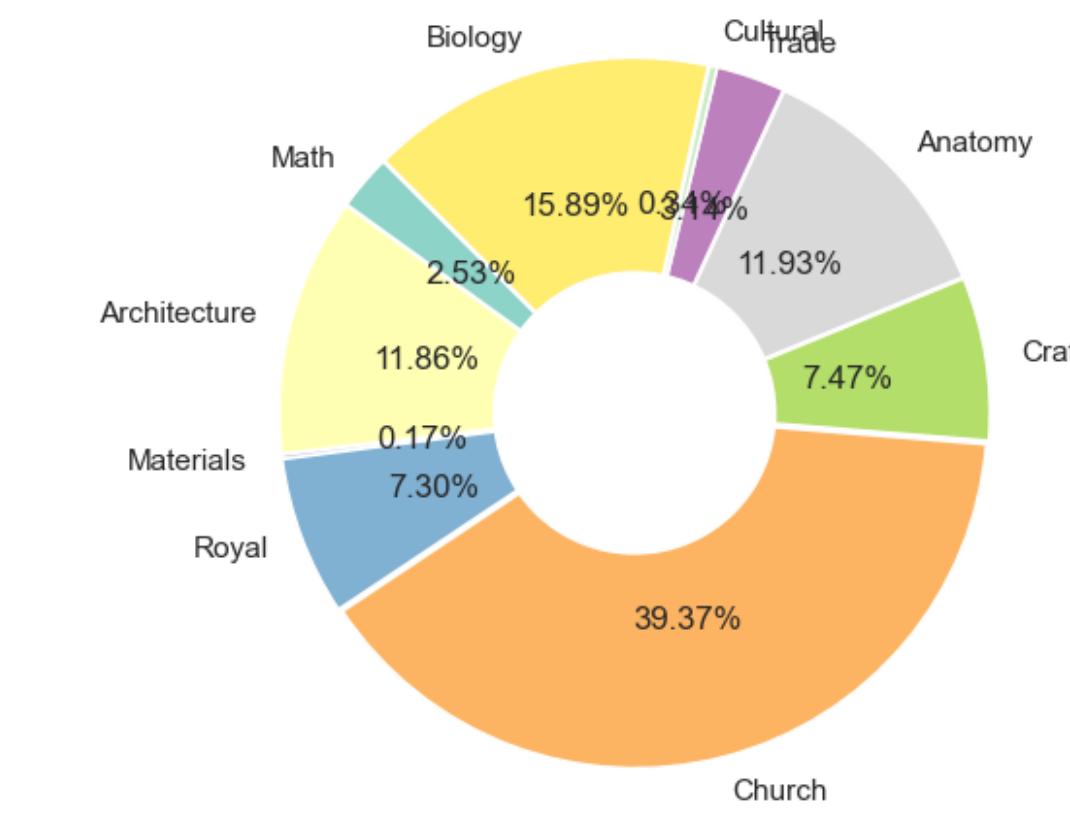
Topic Distribution in Version1



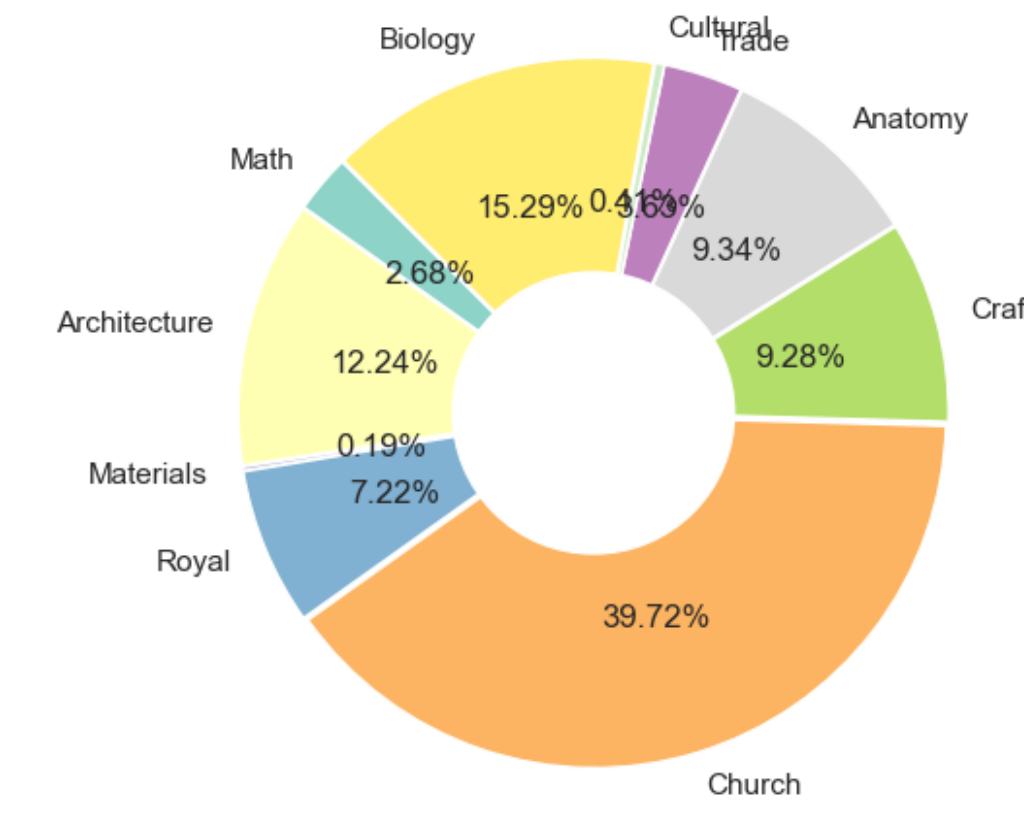
Topic Distribution in Version2



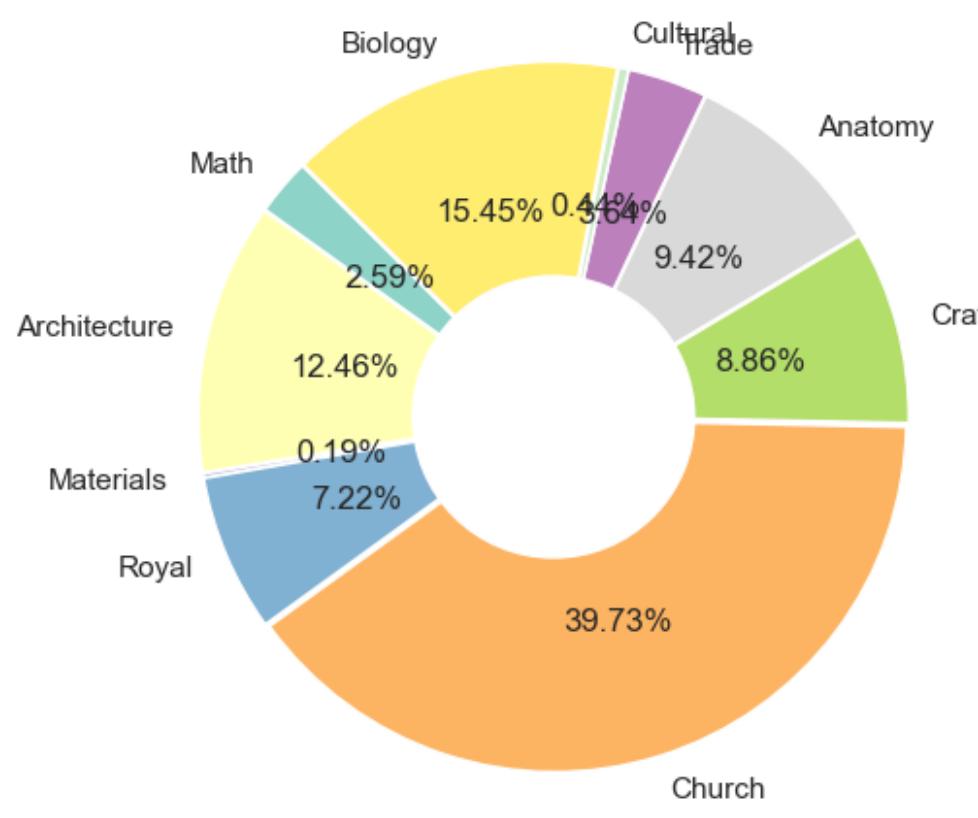
Topic Distribution in Version3



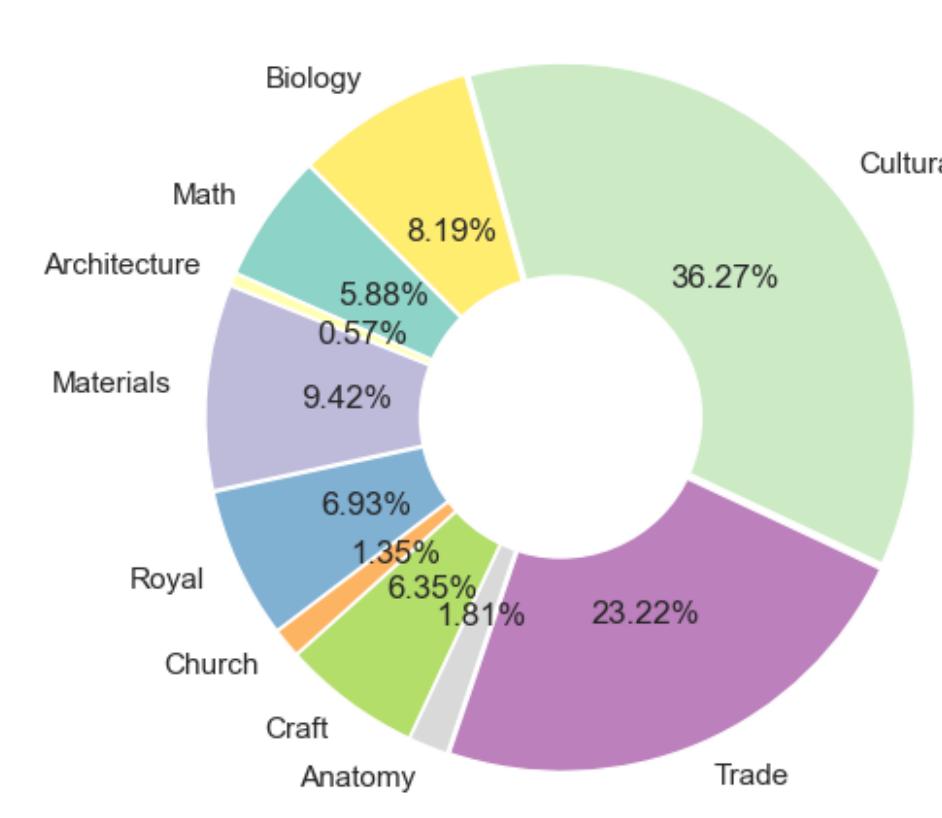
Topic Distribution in Version4



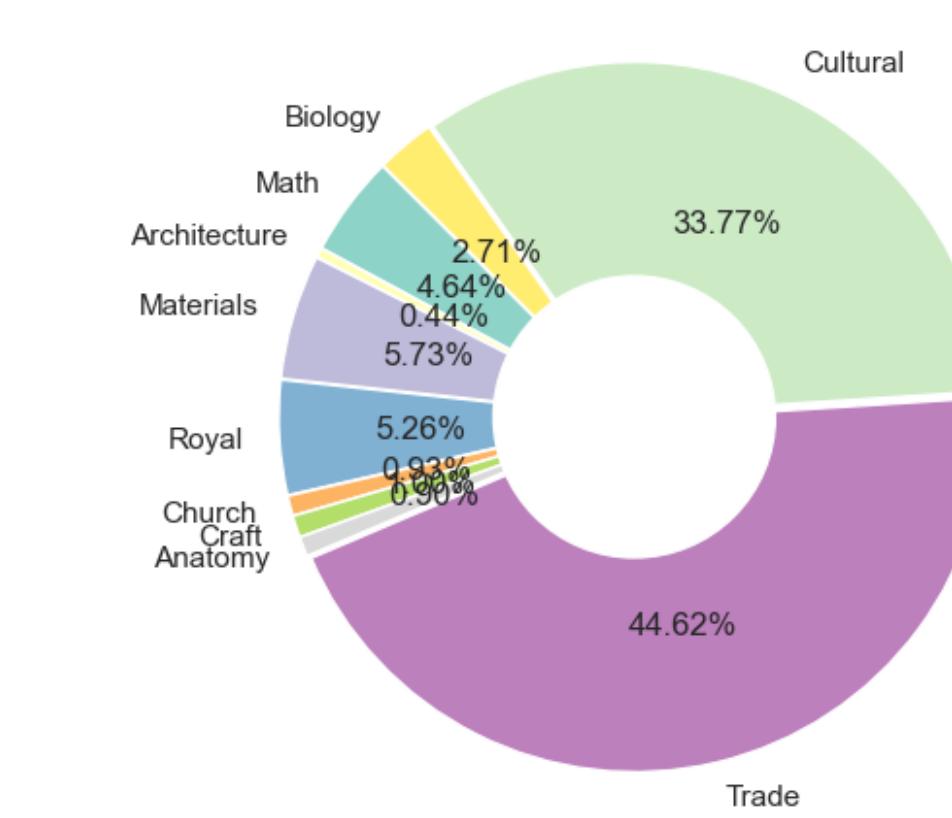
Topic Distribution in Version5



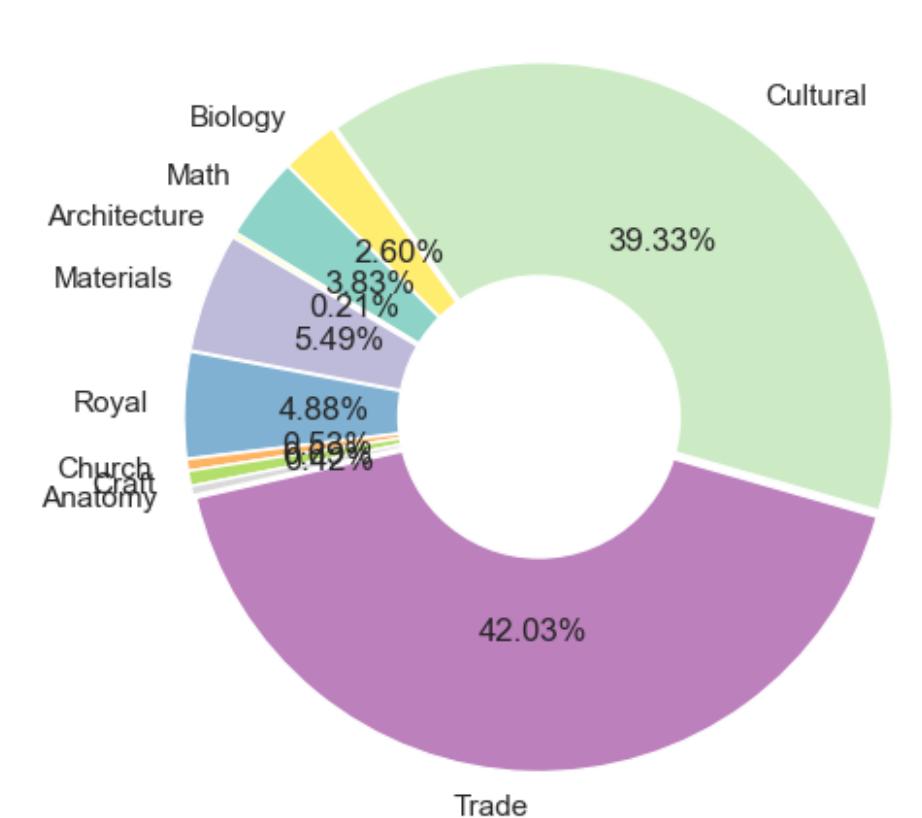
Topic Distribution in Version6



Topic Distribution in Version7

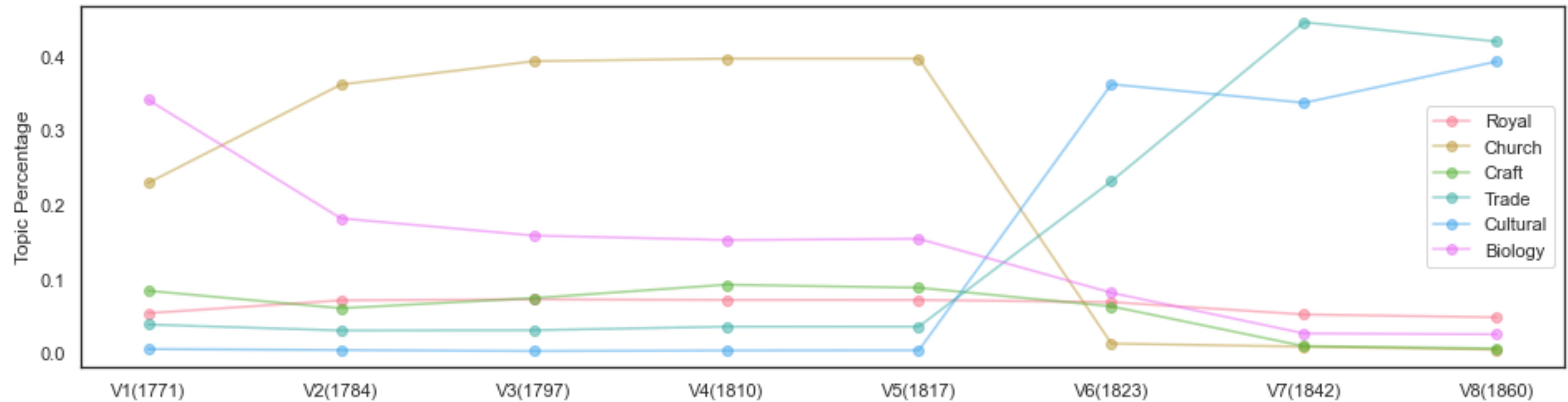


Topic Distribution in Version8



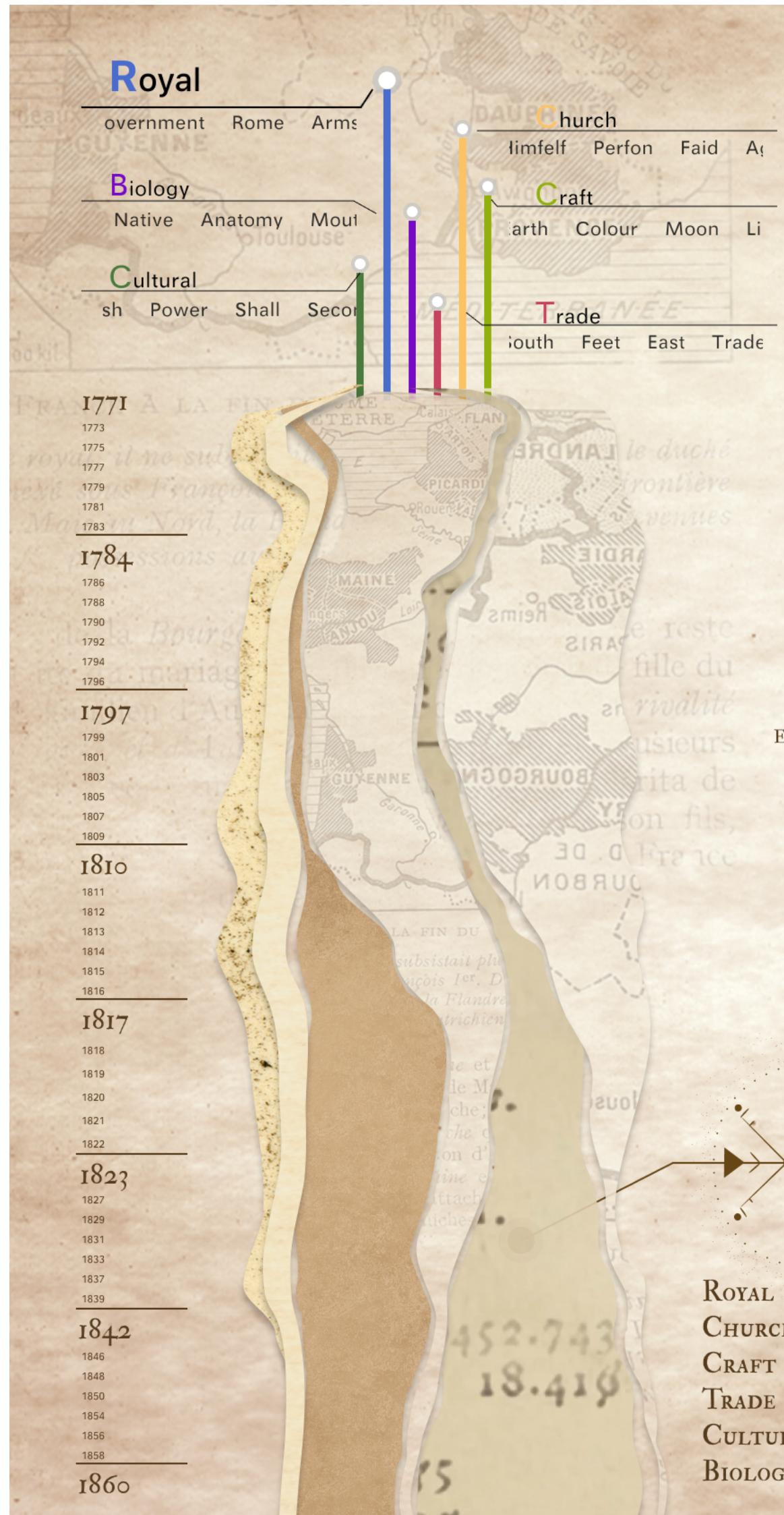
Percentage of Topics per Edition

## #4 Analyze 6 significant topics



发现了？

- 补一张场景图



# Encyclopædia Britannica

— Topics in LDA model —

