Identifying the main objectives of policies in support of business innovation

Draft pilot analysis - May 2022

Background and motivation

- STIP Compass has a wealth of STI policy data, including information on **more than 6500 initiatives** from **57 countries** and the European Union.
- While STIP Compass provides access to this data in various ways (e.g. dashboard, search tool, download), **the portal does not facilitate summarising and benchmarking the types of policies countries have introduced**. As a result, policymakers can be overwhelmed with the amount of information.
- To address this limitation, we need to **explore ways to synthesise country information** and to **identify key patterns in the data**.

Objectives of the pilot

- Using the data reported in the 2021 edition of the STIP survey, **identify the main overarching topics covered by policy initiatives** in the “business innovation and innovative entrepreneurship” policy area. Conducted in collaboration with TIP work in this area.
- Estimate **how much prominence** countries give to these topics in the policies they report to the survey.
- Measure how the prevalence of topics **changes over time**.
- **Visualisation of the prevalence of topics given by countries and over time**, comparing to the OECD median.

Leveraging the 2017, 2019 and 2021 STIP survey datasets: Summary of business innovation policies

- Focus on selected countries with comparatively higher data quality and/or engaged in CSTP / WP projects:
  - Australia, Brazil, Canada, France, Germany, Korea, Italy and the United Kingdom.

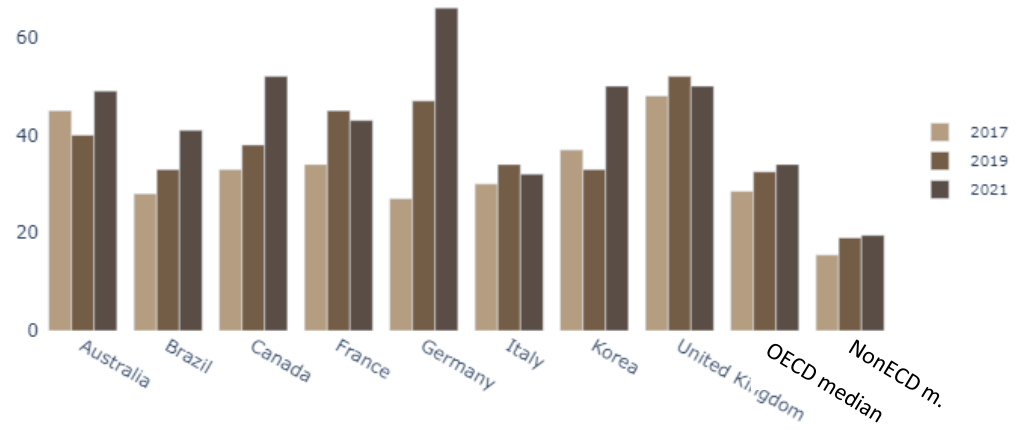
- Text corpus built by aggregating the "**short description**" and "**objectives**" fields per initiative.
- Such a corpus has increased in size over the three survey editions:

Table 1: Description of corpus by survey (OECD countries only)

Survey year	Total words in corpus	Participant countries	Total initiatives
2017	71520	38	1144
2019	102459	38	1281
2021	120166	38	1459

- Distribution of initiatives across the 3 editions of the survey:

Figure 1: Evolution of initiatives dedicated to Innovation in firms and innovative entrepreneurship policy area by selected countries and by survey



Describing the corpus (policy initiatives texts): Are country corpuses' comparable?

- In terms of initiative texts' length:
  - A small set of words used to describe initiatives (increasing over time)
  - Selected countries close to OECD mean words count, except Korea
- In terms of lexical diversity:
  - Large range of lexical diversity amongst countries
  - Selected countries close to OECD median or higher
  - English as official language in country <u>not determining factor</u> of lexical diversity

Figure 2: Average number of words per initiative by selected countries

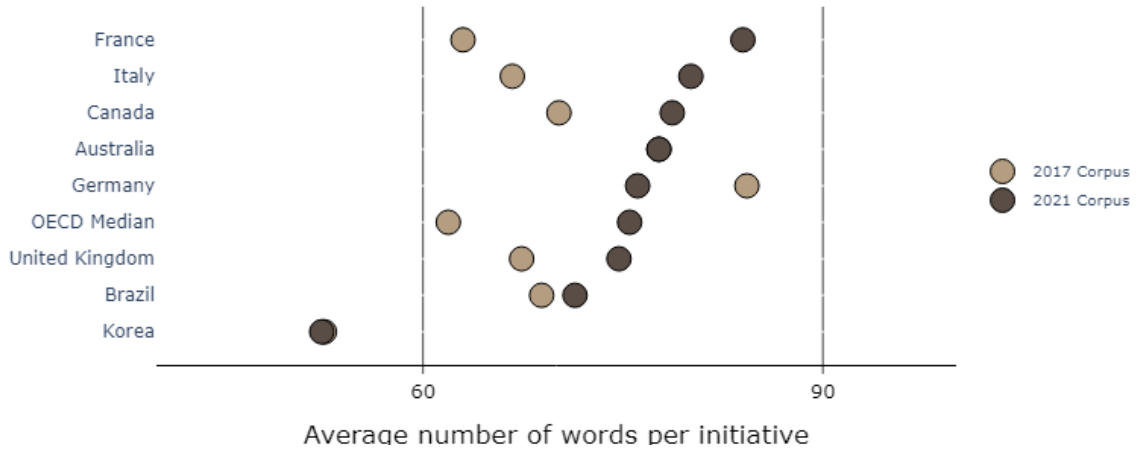
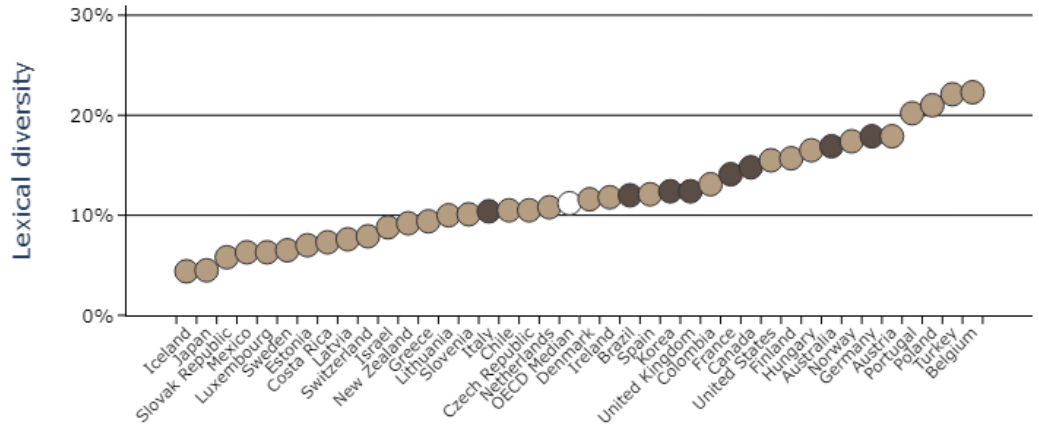


Figure 3: Lexical diversity of total corpus by country



Pre-processing:
How is the data prepared for semantic analysis?

- Replace synonym words/group of words using a vocabulary of concepts in the STI field
- Removing stopwords (including acronyms, locations, adverbs, foreign words)
- Remove shortwords (words below 3 letters)
- Lemmatize (grouping together the inflected forms of a word so they can be analysed as a single item):
  - Remove: determiner, adposition, numeral, symbol, pronoun, conjunction, coordinating conjunction, punctuation, subordinating conjunction, interjection and particle

Using Latent Dirichlet Allocation (LDA) for topic modelling

- Explanation of latent dirichlet allocation (LDA) algorithm
  - The LDA algorithm is a form of unsupervised machine learning commonly used for topic modelling in Natural Language Processing.
  - LDA **explains a set of documents** (policy initiatives, in our case) **in terms of unobserved overarching topics**.
  - A topic is defined as **a set of individual words or combination of words that**, taken together, suggest a relevant theme.
  - Each topic is identified with the presence of **unique combination of key words that tend to appear together** in policy objectives.
  - In sum, applied to our corpus of initiatives, the LDA topic model will **identify emerging topics and calculate how each initiative is associated with one or more of such topics** (topic scores).
- How to define an adequate number of topics? Using the coherence metric and hyperparameters
  - The number of topics the model will use is an input to the LDA model.
  - A common approach to evaluating the most appropriate number of topics is to optimize the notion of coherence (such as Roeder M., Both A. & Hinneburg A., 2015).
  - Coherence increases with the number of topics. Here, optimising coherence means identifying an inflexion point where coherence growth flattens, and which also **provides interpretable topics**.

Topics identification

- 7 topics identified in the full corpus, presented here by order of prevalence in the corpus:

Topic	Salient keywords
1. Leveraging the RDI ecosystem	Research, firm, centre, cooperation, international, scientific, university, technology transfer
2. Access to finance	Firm, fund, startup, investment, enterprise, capital, finance, loan
3. Entrepreneurship and business support services	Entrepreneurship, business, network, information, incubator, training, advice
4. Stimulating demand for innovation	Innovation, firm, procurement, government, public, partnership
5. Tax incentives for R&D and innovation	Firm, investment, incentive, rate, corporate, fiscal, deduction, relief
6. Industrial technology development	Industry, technology, manufacturing, sector, academia, security,
7. Tackling societal challenges	Employment, energy, environmental, social, sustainable, clean, transition, mobility, climate

Robustness check

To verify the model's coherence, we inspect the topic scores that the model assigns to each initiative.

A topic score is a value ranging between [0,1] that indicates the probability for the initiative to belong to a given topic. It can be interpreted as a **measure of how much the initiative's text emphasises a given topic**.

Below are some examples of the highest scoring initiatives for each topic.

Topic	Highest scoring initiative
1. Leveraging the RDI ecosystem	**Open Innovation** (Hungary, 2017 STIP survey) Strengthening cooperation and knowledge flow between SMEs and large companies to increase competitiveness of SMEs.
2. Access to finance	**Slovene enterprise fund (SEF) – Venture Capital** (Slovenia, 2017 STIP survey) Support that includes entry into the ownership structure and enterprise management together with private investors with capital investments (venture capital and mezzanine capital). […]
3. Entrepreneurship and business support services	**Small Business Services** (Canada, 2021 STIP survey) One-stop-shop facilitating access to advisors who help connect entrepreneurs and business owners with available programming.
4. Stimulating demand for innovation	**Procuring innovation initiative** (Ireland, 2019 STIP survey) Encourages procuring authorities to adopt a more open approach to procuring goods and services [by seeking solutions in the market-place […] Reviews and updates the Government's Procurement guidelines and procedures obstacles to SMEs […]
5. Tax incentives for R&D and innovation	**CIR-CIR R&D Tax Credit** (France, 2021 STIP survey) A measure to support the research and development activities of companies, with no sector or size restrictions. Companies that spend on basic research, applied research and experimental development can benefit from the CIR. […]
6. Industrial technology development	**Next Generation Technologies Fund** (Australia, 2021 STIP survey) An investment of $730 million spanning 2016-2026 aiming to accelerate R&D in technologies for the "future Defence force after next".
7. Tackling societal challenges	**ENOVA** (Norway, 2021 STIP survey) A public enterprise promoting a transition to environmentally sustainable energy production and use, and the development of energy and climate technology.

Calculating topic prevalence by country

- For a given topic,
  (1) We calculate the absolute topic score for country X in year Y by adding initiatives' individual scores. We identify the OECD median score for each topic.
  (2) We normalise the values in (1) by the number of words reported by the country X in year Y. We do the same for the OECD median topic score.
    - We normalise calculations by country to account for the differences in reporting highlighted by previous slides.
  (3) The normalised topic scores can be divided by the OECD median topic score. This allows comparing how prevalent topics are in countries relatively to the OECD median.

- In theory, the above calculation could also be weighted by initiatives' budgets. However, as we attempted this approach we have identified important inconsistencies in how countries report budget data. Such inconsistencies propagate into the results, making them unreliable.

➢ As topic prevalence is not weighted by budget, **this metric should not be interpreted a measure of policy effort**.

➢ Rather, **topic prevalence indicates how more or less frequently countries raise a given topic in their policies** (relative to others). <u>It does not intend to evaluate performance, but rather to help characterise the policy mix.</u>

Topic prevalence in corpus

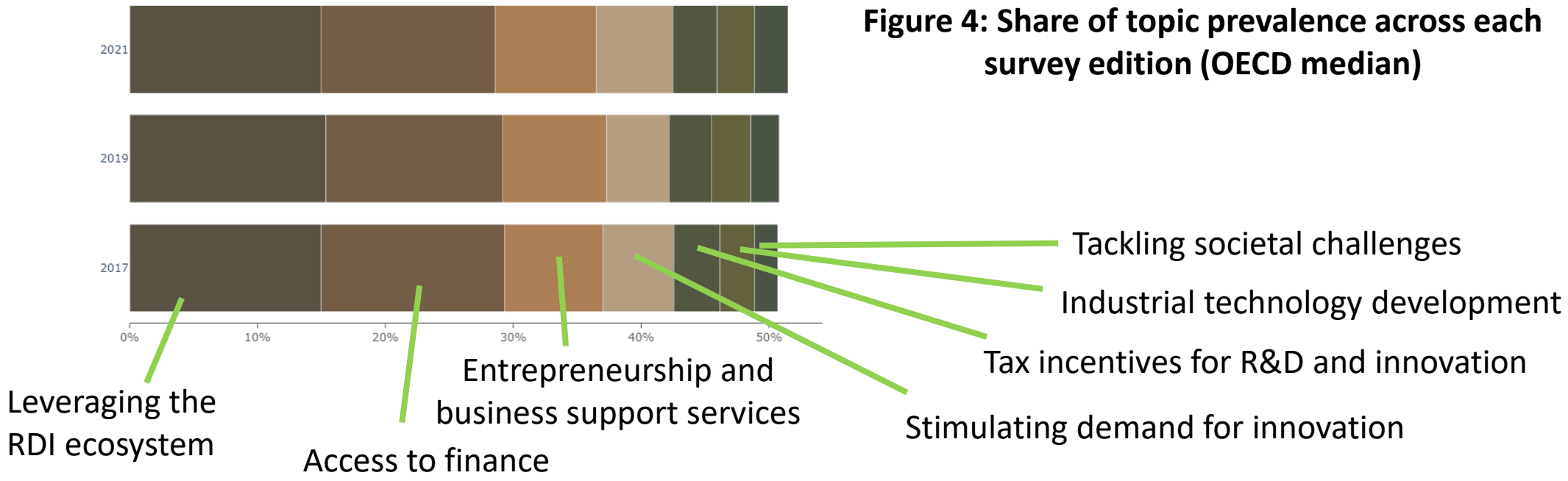


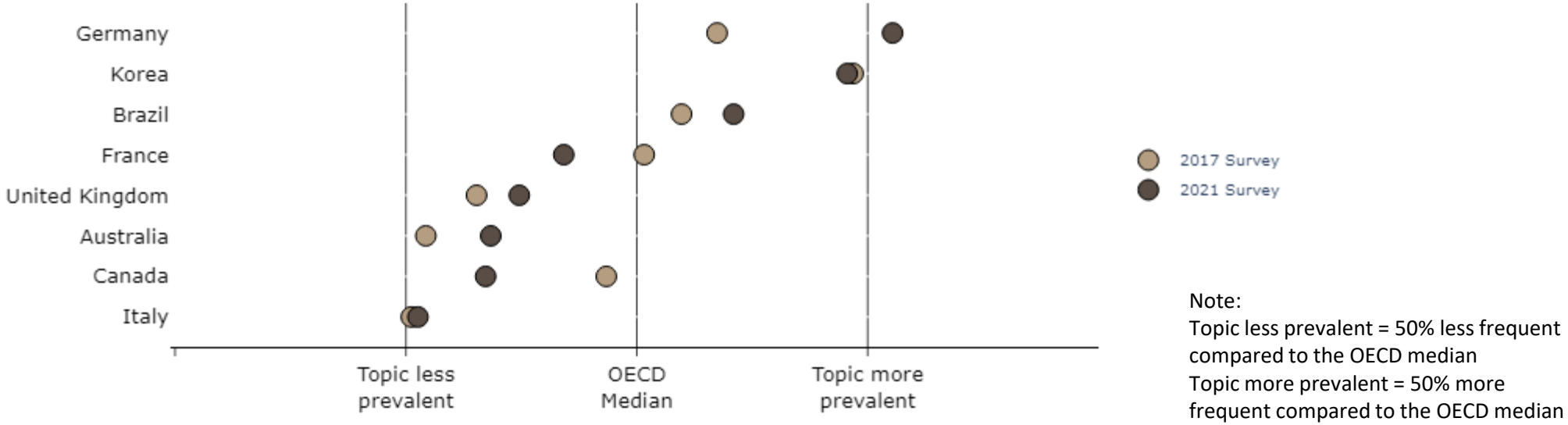
Figure 4: Share of topic prevalence across each survey edition (OECD median)

- About 50% of the text in policy initiatives is about one of the topics retained from the model.
- The model identified other topics, that were not retained for the analysis:
  - Too vague, e.g. business innovation in general (16% topic prevalence) or generic policy support (5%)
  - Not significant for our purposes, e.g. generic national strategies (15%)
  - Apparently spurious (no coherence among keywords): 13%
  - Health care, a small topic addressed by a few initiatives: (1% topic prevalence)

Preliminary results by topic

1. Leveraging the RDI ecosystem

Figure 5: Prevalence of topic in countries' declared objectives vs OECD Median

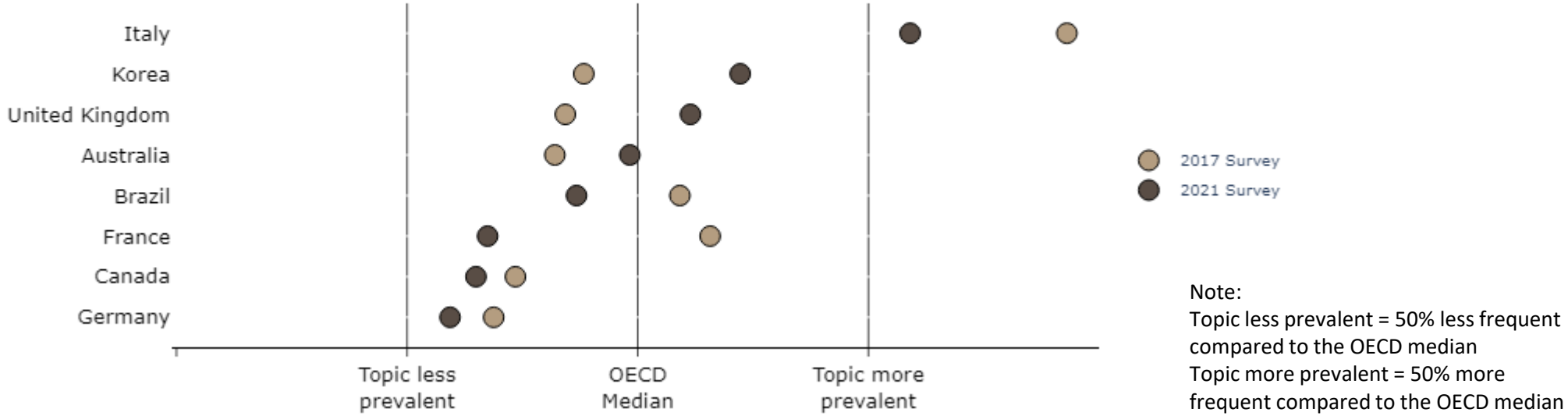


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- Compared to other countries, **German** and **Korean** business innovation policies most frequently emphasise linkages with the RDI ecosystem, followed by **Brazil**. In **Germany**, emphasis in 2021 has increased markedly vis-à-vis 2017.
- In other countries, topic prevalence is lower than OECD median, with
  - **Italy** emphasising the topic the least,
  - **Canada** and **France** emphasising the topic less in 2021 compared to 2017.

2. Access to finance

Figure 6: Prevalence of topic in countries' declared objectives vs OECD Median

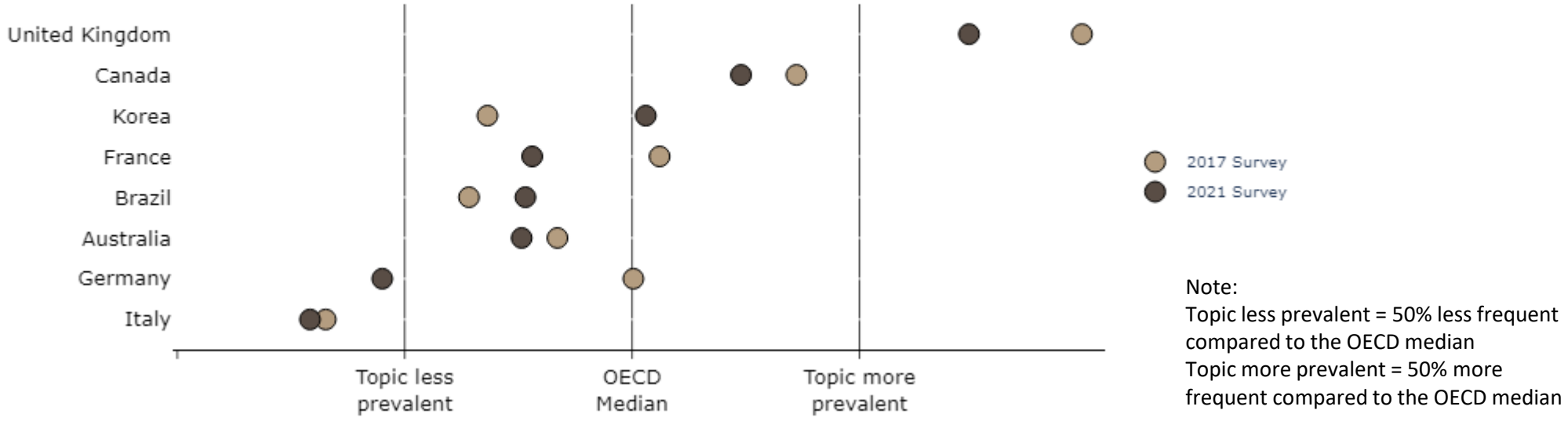


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- Among the selected countries, the topic is most prevalent in the policies reported by **Italy**
- This topic has been emphasises with increased frequency in the policies reported by **Korea** and the **United Kingdom**.
- By contrast, policies in **France** and (to a lesser extent) in **Brazil** shifted away from this topic in the 2021 edition of the survey.
- Topic prevalence for **Canada** and **Germany** remained stable (and low)**.**

3. Entrepreneurship and business support services

Figure 7: Prevalence of topic in countries' declared objectives vs OECD Median

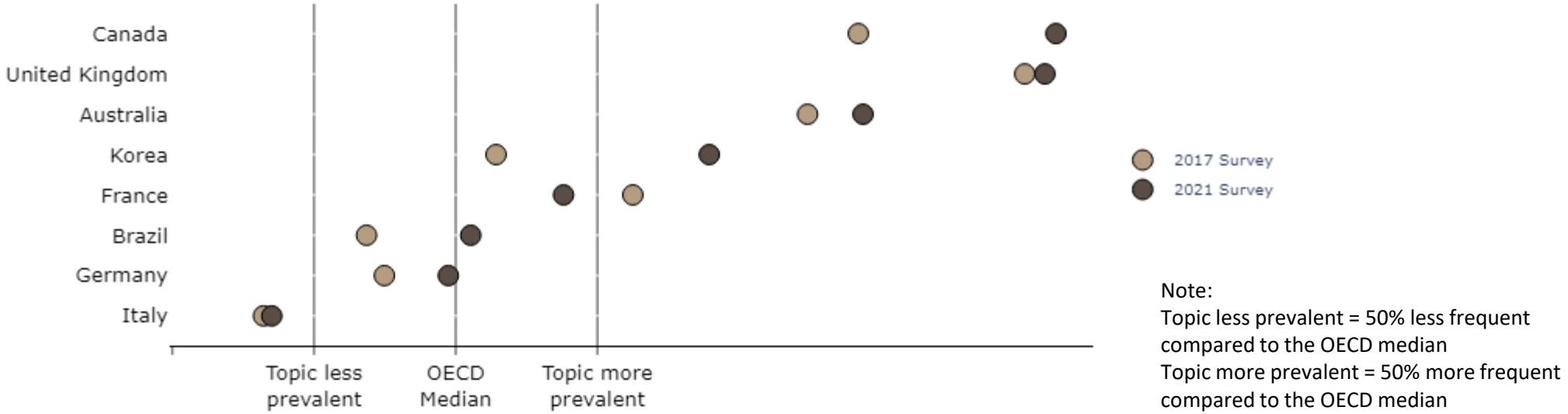


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- Topic is most prevalent in **British** and **Canadian** policies, with **Korean** policies reaching the median in 2021.
- It is least prevalent in **Italian** and **German** data (for the latter, it was around the median in 2017).
- Policies in other countries (**France**, **Brazil** and **Australia**) tackle entrepreneurship and business support services slightly below the OECD median.

4. Stimulating demand for innovation

Figure 8: Prevalence of topic in countries' declared objectives vs OECD Median

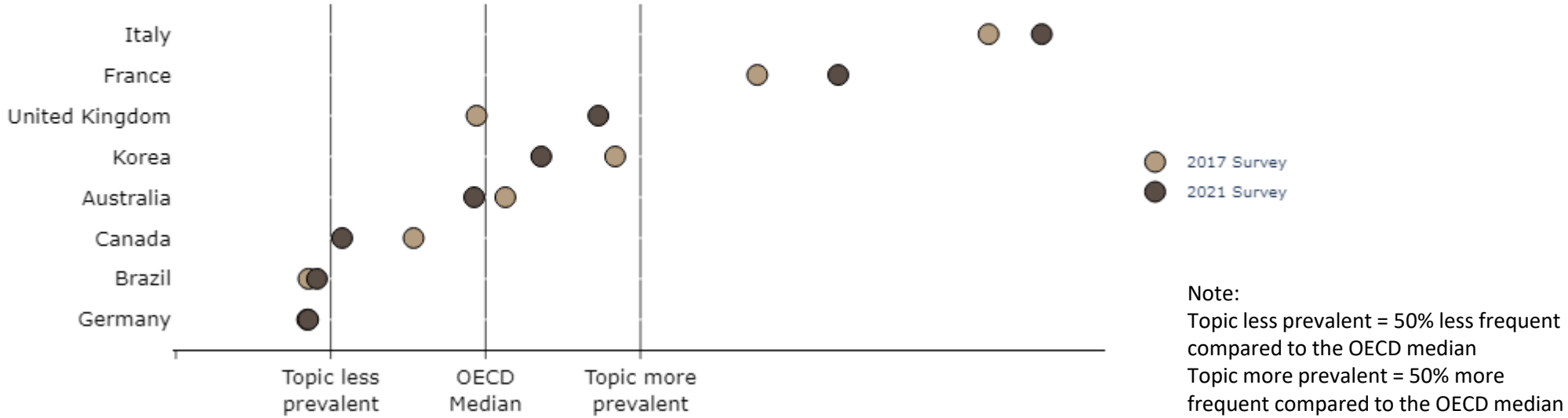


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- In the 2021 survey response, more policies in **Canada** emphasise stimulating demand for innovation. While smaller in amplitude, there has also been an increase in policies reported by **Korea** (high prevalence) and **Brazil** (close to OECD median).
- **Australia** and the **UK** also stand out for emphasising procurement and related activities.
- By contrast, topic prevalence is low in **Italy**.

5. Tax incentives for R&D & innovation

Figure 9: Prevalence of topic in countries' declared objectives vs OECD Median

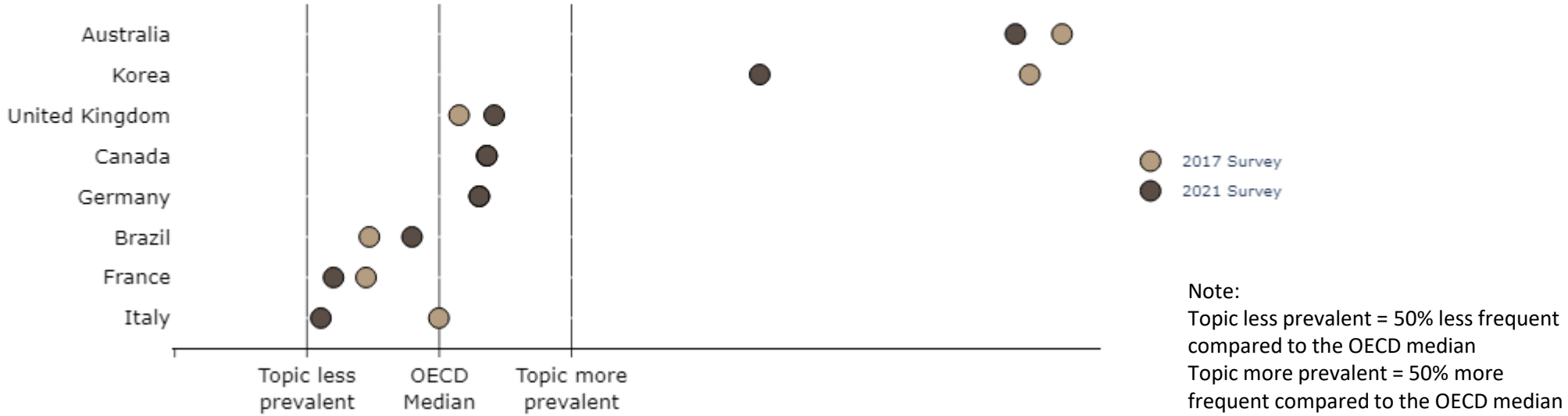


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- **Italy**, **France** and (to a lesser extent) **UK** reported policies stand out by emphasising this topic most recurrently to other countries (with an increase in 2021).
- Topic prevalence is close to the OECD median in **Korea** and **Australia**.
- Among the selected countries, the topic is least prevalent in **Germany**, **Brazil** and **Canada**.

6. Industrial technology development

Figure 10: Prevalence of topic in countries' declared objectives vs OECD Median

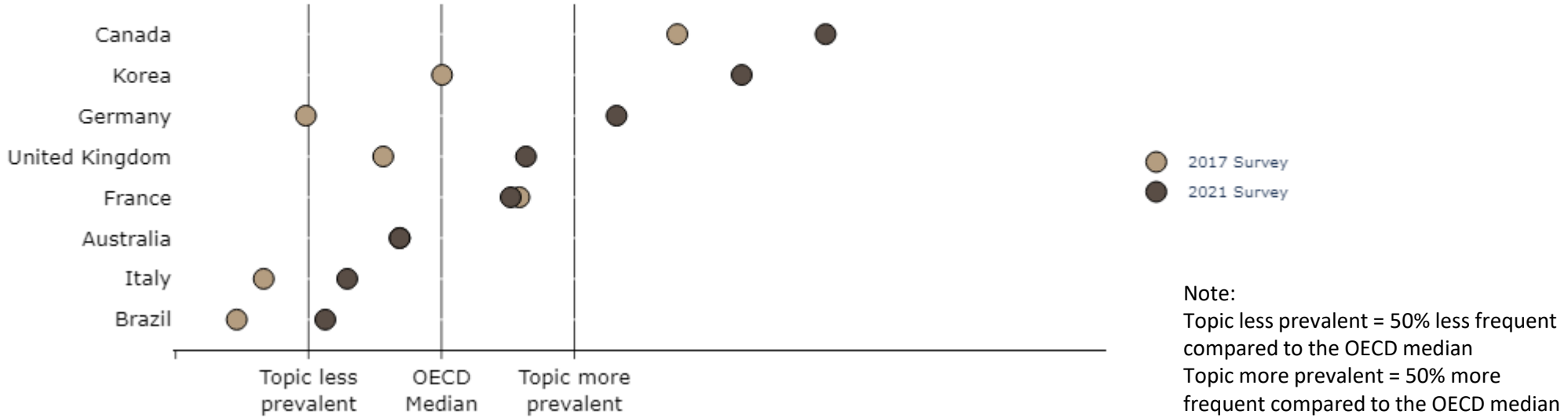


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- **Australian** and **Korean** policies stand out for emphasising industrial technology development.
- In **France** and **Italy**, the topic is less prevalent (decreasing in the latter relative to 2017).
- Other countries are around the mean (and stable).

7. Tackling societal challenges

Figure 11: Prevalence of topic in countries' declared objectives vs OECD Median



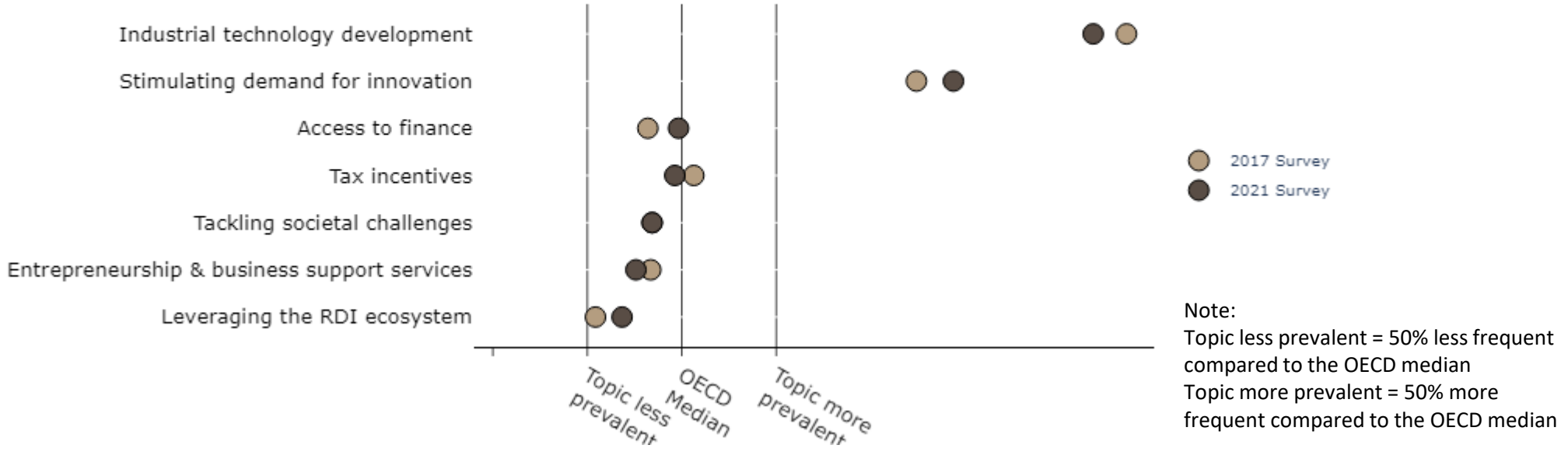
Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- Societal challenges have become very prevalent in the policies reported by **Canada**, **Korea** and **Germany**. Prevalence of this topic has also grown in the **United Kingdom**. It has remained higher than the OECD median in **France**.
- Compared to 2017, the policies reported by **Brazil** and **Italy** increasingly emphasise societal challenges (though remain below the OECD median).

Preliminary results by country

Australia

Figure 12: Prevalence of topics in Australia's declared objectives vs OECD Median

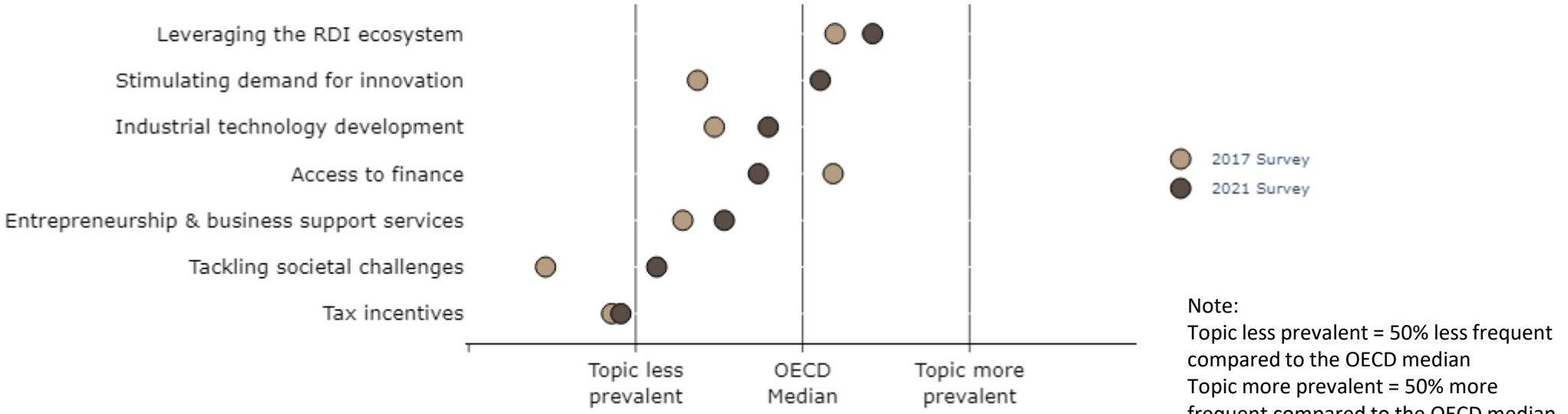


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- Australia's policies stands out for emphasising **industrial technology development** and **stimulating demand through innovation**.
- Otherwise, most of Australia's topics' prevalence in 2021 follow closely those of the OECD median. Relative to other topics, **leveraging the RDI ecosystem** is the least frequent topic.

Brazil

Figure 13: Prevalence of topics in Brazil's declared objectives vs OECD Median

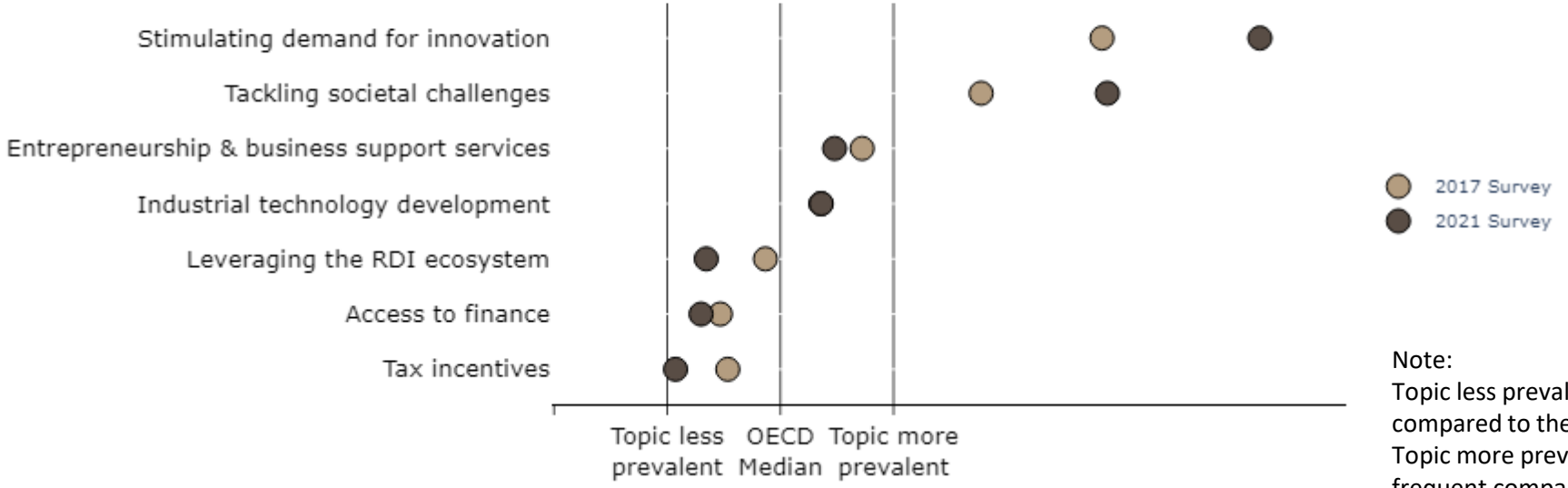


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- A lower share of the text data reported in Brazilian policies can be associated with the selected topics than the OECD median (43% vs 53% in 2021). It may be that Brazilian policies tackle a **distinct set of topics** beyond those identified in by the model.
- The exceptions are **leveraging the RDI ecosystem** and **stimulating demand for innovation**, which are more prevalent than OECD median.
- Relative to 2017, the policies reported in 2021 further emphasise **stimulating demand for innovation** and **tackling societal challenges**.

Canada

Figure 14: Prevalence of topics in Canada's declared objectives vs OECD Median

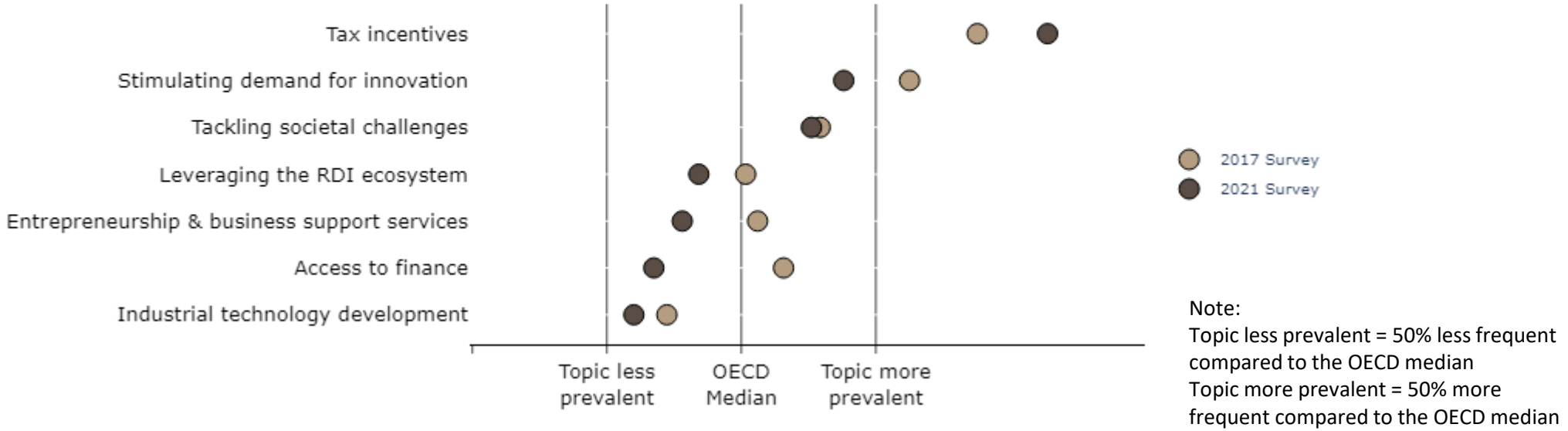


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- Canadian policies stand out for emphasising **stimulating demand through innovation** and **tackling societal challenges** (with additional emphasis in 2021).
- The topics of **tax incentives**, **access to finance** and **leveraging the RDI ecosystem** come up less frequently compared to the OECD median.

France

Figure 15: Prevalence of topics in France's declared objectives vs OECD Median

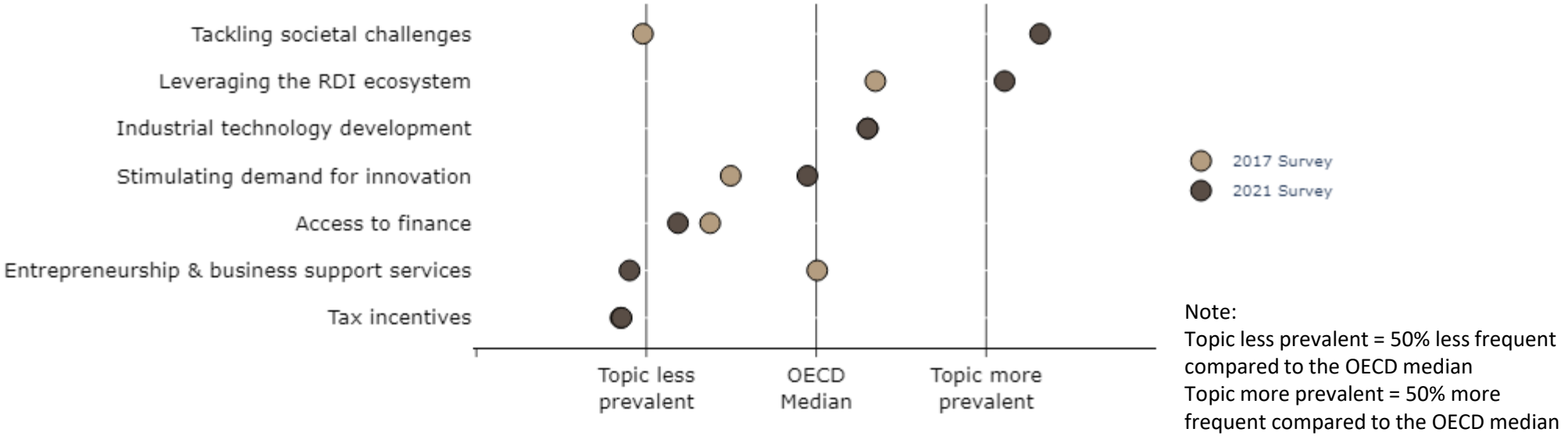


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- Compared to some of the other countries in this analysis, France's business innovation policy mix is more distinct relative to the OECD median.
- By far the most prevalent topic in France is **tax incentives,** followed by **stimulating demand through innovation** and **tackling societal challenges**.
- The least prevalent is **industrial technology development**.
- Relative to 2017, French policies reported in the 2021 survey emphasise less the topics of **access to finance**.

Germany

Figure 16: Prevalence of topics in Germany's declared objectives vs OECD Median

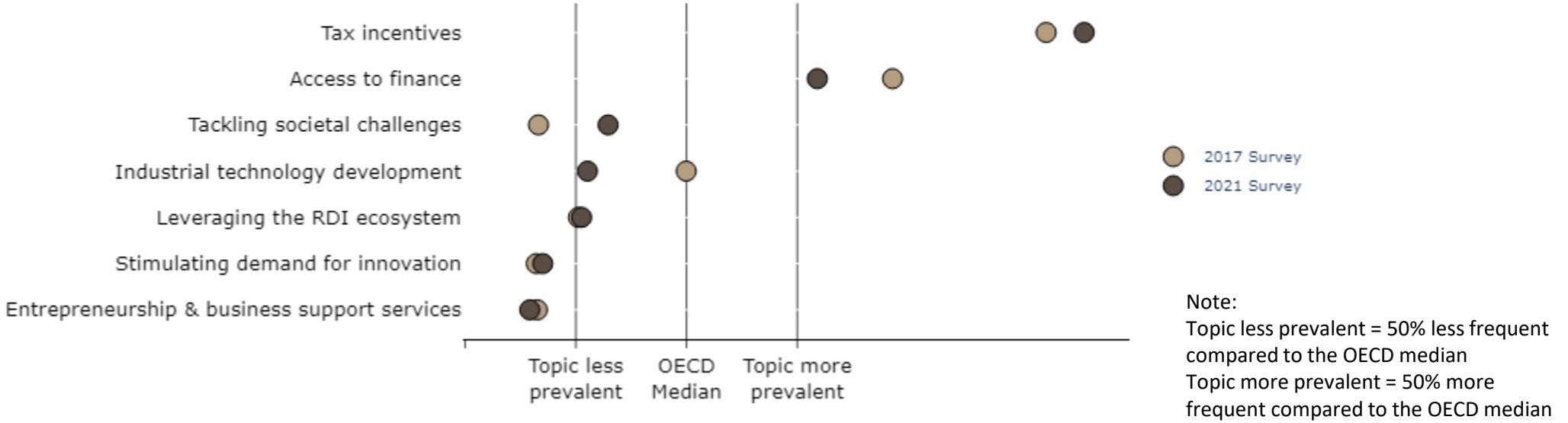


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- As in France, Germany displays a distinct policy mix relative to the OECD median.
- Policies place emphasis in **tackling societal challenges** and **leveraging the RDI ecosystem**. Such emphasis has increased in the 2021 edition of the STIP survey, particularly for the former.
- By contrast, business innovation policies reported by Germany emphasise relatively less frequently the topics of **tax incentives**, **entrepreneurship and business support services** and **access to finance**.

Italy

Figure 17: Prevalence of topics in Italy's declared objectives vs OECD Median

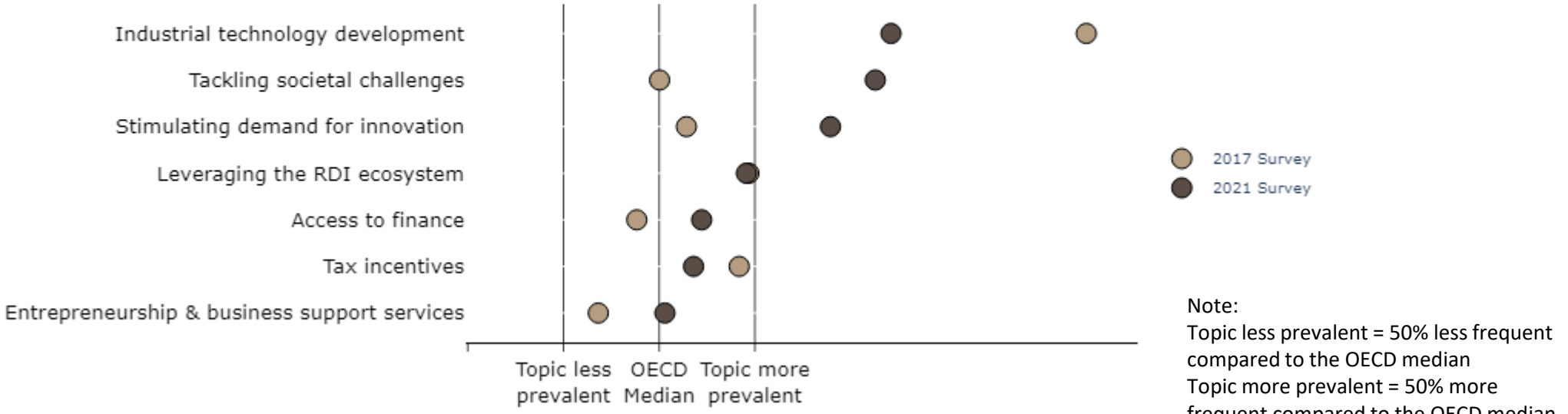


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- Most topic frequencies in Italy diverge from the OECD median, also revealing a distinct policy mix.
- Italy's policy initiatives most recurrently emphasise **tax incentives** and **access to finance**.
- In 2021, Italy increased the mentions of **societal challenges** in its business innovation policies. The opposite trend is observed for **industrial technology development**.
- Other topic frequencies have remained stable in the 2017 and 2021 editions of the survey.

Korea

Figure 18: Prevalence of topics in Korea's declared objectives vs OECD Median

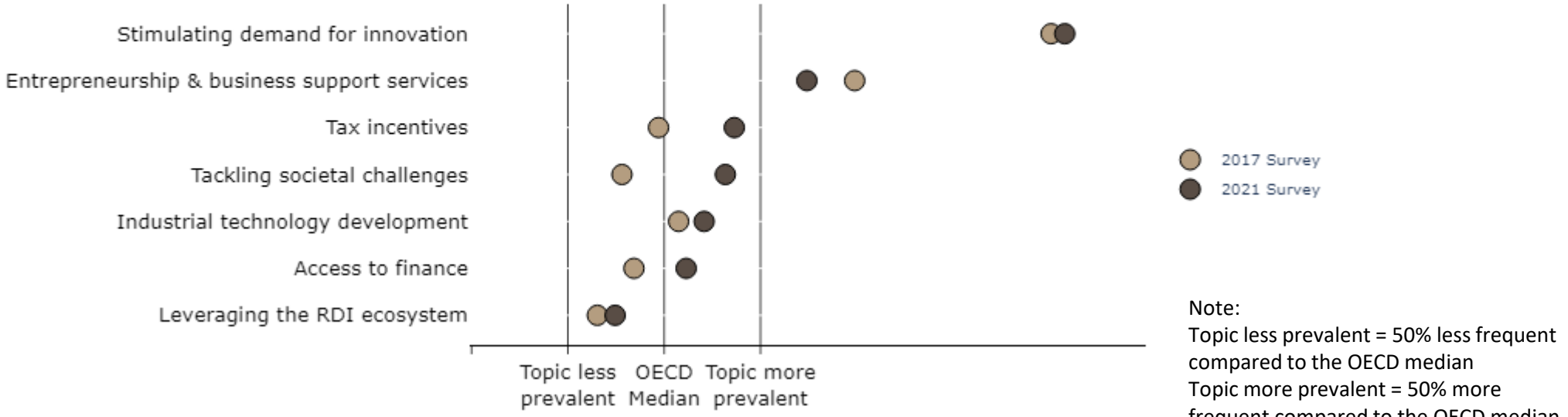


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- In Korea, most topic frequencies have changed significantly between 2017 and 2021.
- In 2021, all topics are more prevalent than OECD median, with **industrial technology development** being the topic most emphasised by Korean policies.
- Mentions of **societal challenges** have increased further beyond the OECD median. Recent policies increased the emphasis on **stimulating demand through innovation**.

United Kingdom

Figure 19: Prevalence of topics in the UK's declared objectives vs OECD Median

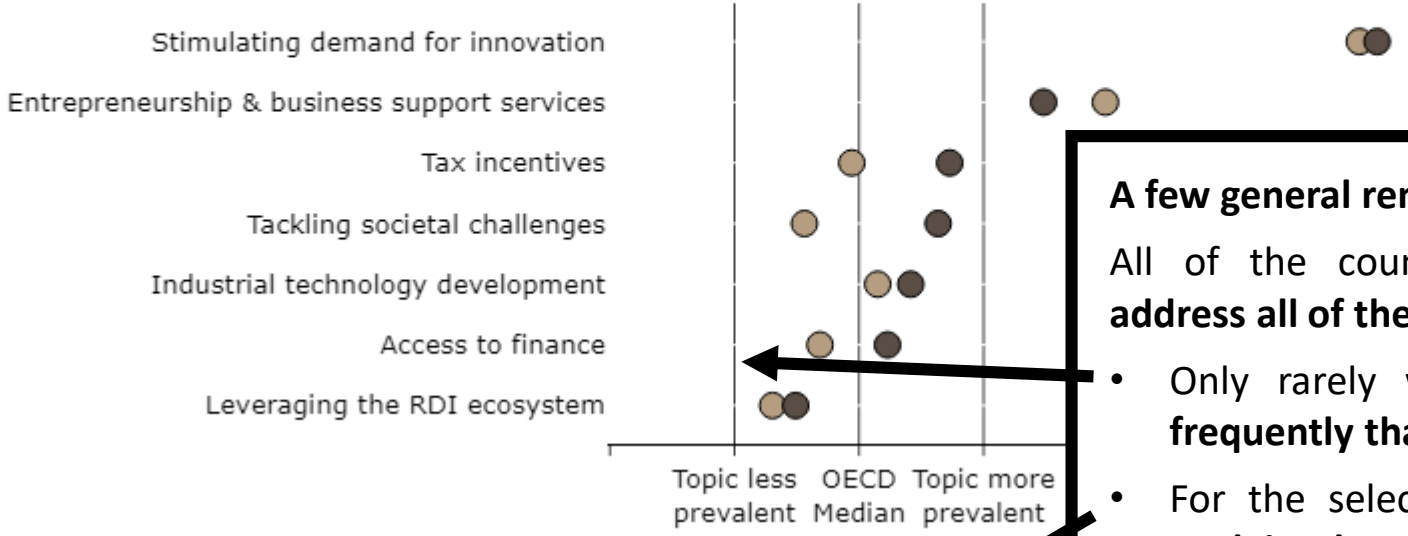


Note:
Topic less prevalent = 50% less frequent compared to the OECD median
Topic more prevalent = 50% more frequent compared to the OECD median

- A higher share of the text data reported in British policies can be associated with the selected topics (64% vs 53% for OECD median in 2021).
- British policies stand out for frequently raising **stimulating demand for innovation**, followed by **entrepreneurship and business support services**, **tax incentives** and **tackling societal challenges**.
- Relative to the OECD median, British policies less frequently mention the topic of **leveraging the RDI ecosystem**.

United Kingdom

Figure 19: Prevalence of topics in the UK's declared objectives vs OECD Median



- A higher share of the text data reported in British policies car
  topics (64% vs 53% for OECD median in 2021).
- British policies stand out for frequently raising **stimulating de**
  entrepreneurship and business support services, **tax incenti**
- Relative to the OECD median, British policies less frequently
  RDI ecosystem.

A few general remarks:

All of the countries included in this analysis **address all of the seven topics**.

- Only rarely we see a topic come up **less frequently than 50% of the OECD median**.
- For the selected countries, the **seven topics explain about half of the text data** (similar to the OECD median in slide 11). The exceptions are Brazil (less emphasis in topics) and the UK (more emphasis in topics).
- Policies reported by countries **vary in the topics they emphasise**.
- Such emphasis can **change from one edition to the survey** to the next.

Comparing semantic analysis results with counting initiatives by policy theme

In the STIP survey, policies are associated to a number of questions or policy themes

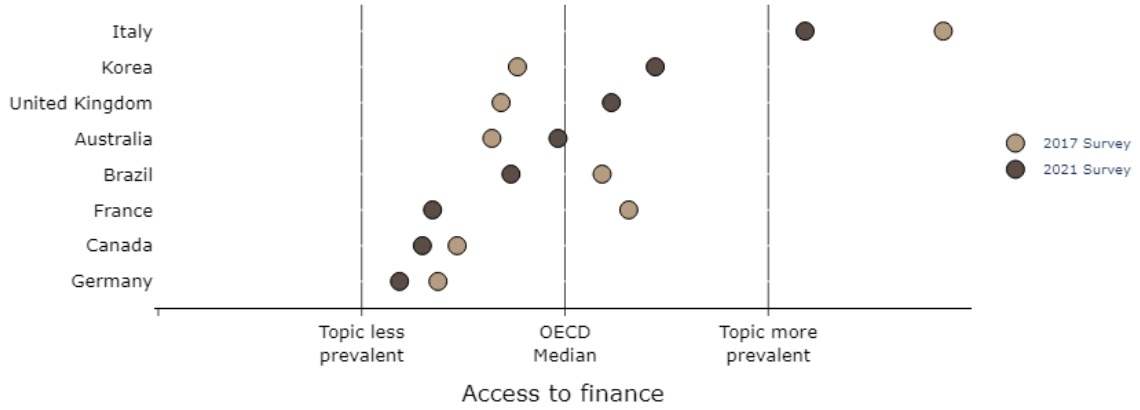
As shown in the table below, there is some overlap between the topics identified by the semantic analysis and the STIP survey themes (questions).

However, **most of the topics emerging from the analysis do not correspond to pre-defined themes in the survey**. In other words, the semantic analysis identifies patterns emerging from the data that go beyond the survey structure.

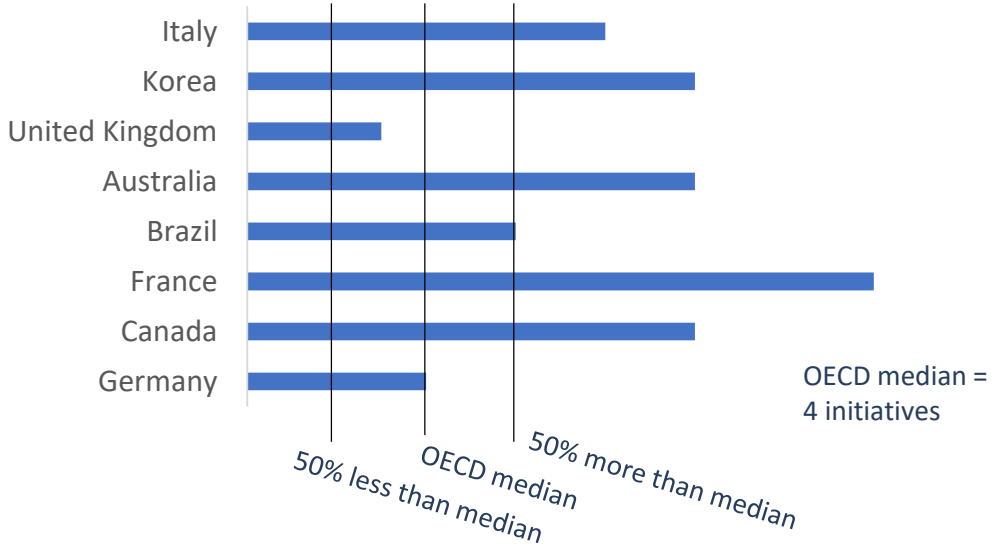
Topic (semantic analysis)	Policy theme (2021 STIP survey)
1. Leveraging the research system	N.A.
2. Access to finance	Access to finance
3. Entrepreneurship and business support services	Partial overlap with “Entrepreneurship capabilities and culture”, “Targeted support to young innovative enterprises” and “Non-financial support for business R&D and innovation
4. Stimulating demand for innovation	Stimulating demand for innovation and market creation
5. Tax incentives for R&D and innovation	N.A.
6. Industrial technology development	N.A.
7. Tackling societal challenges	N.A.

Comparing semantic analysis vs counting numbers of initiatives: Access to finance

Figure 6: Prevalence of topic in countries' declared objectives vs OECD Median



Number of initiatives reported in STIP theme vs OECD Median, 2021



- **Results are very different**. The semantic analysis measures how much the topic is raised **across all initiatives** reported in the business innovation policy area, whereas the initiative count **relies on initiatives tagged under the theme** (question in the survey).
- Initiatives often deal with access to finance but are **primarily addressed at other themes** in the survey policy area (e.g. targeted support for SMEs, national strategies). As a result, while the main initiatives are tagged under a given policy theme, that theme **does not capture how all reported policies emphasise a given topic** in the text.

Concluding remarks and next steps

- Semantic analysis provides a way of **analysing the large volumes of unstructured data** reported in the STIP databases.
  - 350k words on policy objectives around support for business innovation.

- This approach aims to **systematically summarise textual data**, allow **cross-country comparisons** and to observe **shifts in policy trends over time**.
  - It identifies key topics and measures how frequently countries raise them in their policies, giving a sense of orientation of national policy mixes without measuring policy effort. Determining the latter would require additional harmonisation of the budget data reporting in the STIP survey.
- While the robustness checks suggest that the model is coherently clustering initiatives under salient topics, we require further validation.
  - STIP survey national contact points and OECD CSTP country delegates could **assess how accurately the model represents their data** (and the reality of their national business innovation policy). Do data points and trends presented in the pilot **align with how their perceive their business innovation policy mix**?