

NDL デジタルライブラリーカフェ

NDL × データ可視化

2017年11月29日

矢崎 裕一

<http://notation.co.jp>

# 自己紹介：矢崎裕一

職業：データ・ビジュアライゼーション実務家／  
インターフェイス・デザイナー／おもちゃコンサルタント

コミュニティ：コード・フォー・トウキョウ 代表／  
データ・ビジュアライゼーション・ジャパン 発起人

組織：合同会社ノーテーション 代表社員／国立国会図書館 委嘱研究員／多摩美術大学非常勤講師／東京大学  
空間情報科学研究センター 所属

# RESASの教科書

1章 RESASで発見できること

2章 RESASマップ総覧

3章 自治体・地域の分析事例集

4章 Excelで簡単にできる統計解析

…おたに 代表取締役 小谷 祐一朗氏

5章 教育現場における活用の実際

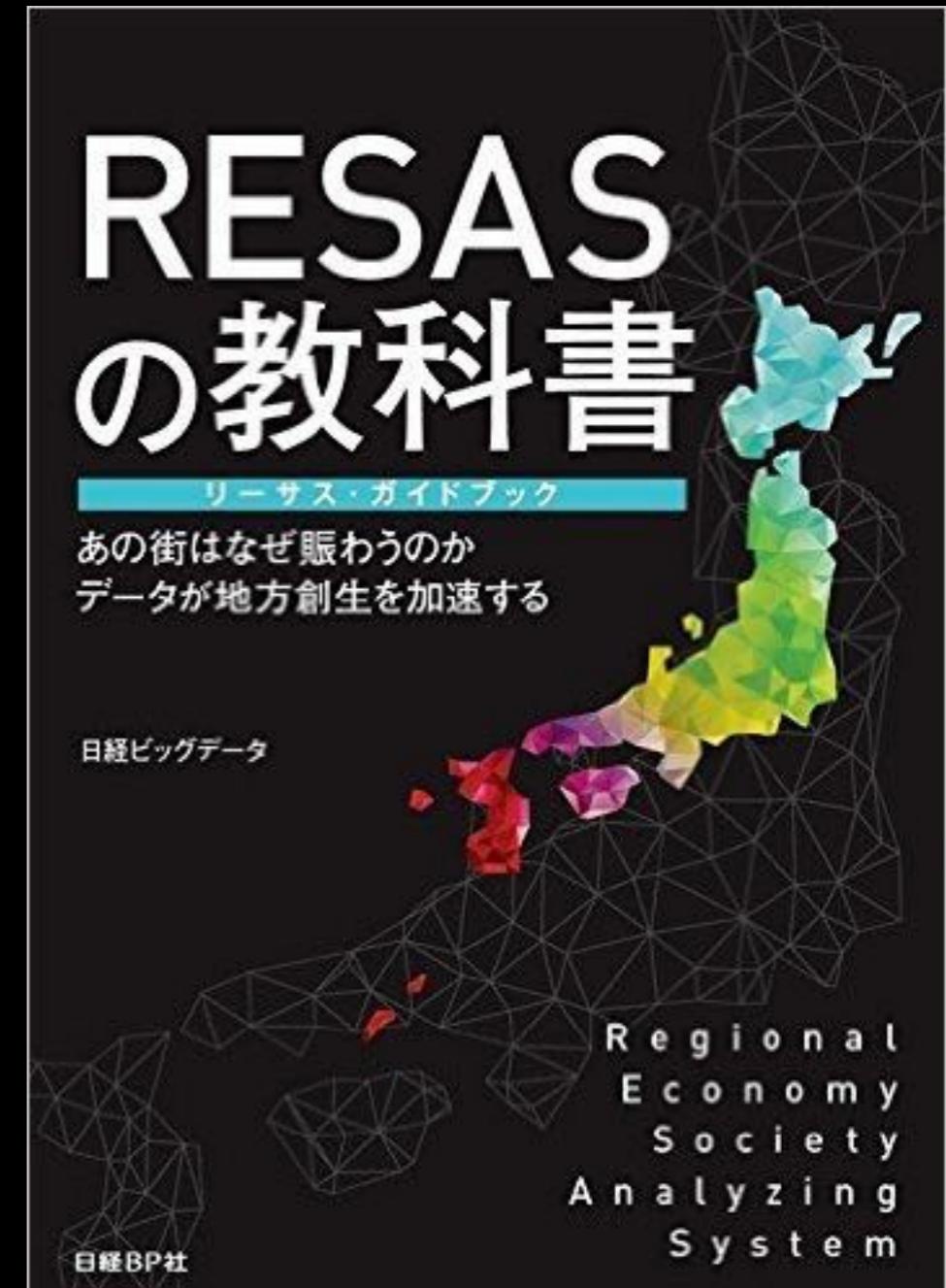
…金沢大学 松浦 義昭氏

6章 地元でワークショップを開こう

…Code for Tokyo 矢崎 裕一&榎本 真美

7章 進化を続けるRESAS

8章 活用を加速するオープンデータ集



<http://www.amazon.co.jp/dp/4822236609>

キーワードを入力

ニュース  

矢崎裕一

藤井一貴



トップ

速報

写真

動画

雑誌

個人

ビジネス

特集

意証調査

ランキング

新着記事一覧

国内

国際

経済

エンタメ

スポーツ

IT・科学

ライフ

オーサー一覧

## 矢崎裕一 新しい単位

YUICHI YAZAKI



矢崎裕一

データ・ビジュアライゼーション実験  
室

yuichi.yazaki

official site

Notation LLC

コード・フォー・トウキョウ 代表/データ・ビジュアライゼーション・ジャパン 発起人/合同会社ノーテーション 代表社員/東京大学空間情報科学研究センター 教員/おもちゃニンサルタント。株式会社ビジネス・アーキテクツにてデザイナー及びアートディレクターを7年間経験後、2008年に独立。独立後数年は、パソコン以外のデバイス（デジタルカメラ、カーナビ、デジタルテレビのインターフェース等）のユーザー・インターフェイスに特化した受託活動を行っていた。近年からは、データ・ビジュアライゼーションの実践と普及に関する様々な活動をおこなっている。

記事

1

オーサーコメント

0

記事一覧

1~1/1件 年月で絞り込む

社会課題を可視化する「データ・ビジュアライゼーション」とは

# Yahoo!ニュース個人で社会課題についての記事を連載開始

<http://bylines.news.yahoo.co.jp/yazakiyuichi/>

31 MAY GALA DINNER IN LISBON

573 PROJECTS SUBMITTED

51 COUNTRIES REPRESENTED

355 SLACK CHANNEL MEMBERS

# THE DATA JOURNALISM AWARDS 2018

Deadline: 26 March 2018 23:59 GMT

APPLY NOW

Save the date



12 WINNERS

GEN SUMMIT 2018

1951 NEWSLETTER SUBSCRIBERS

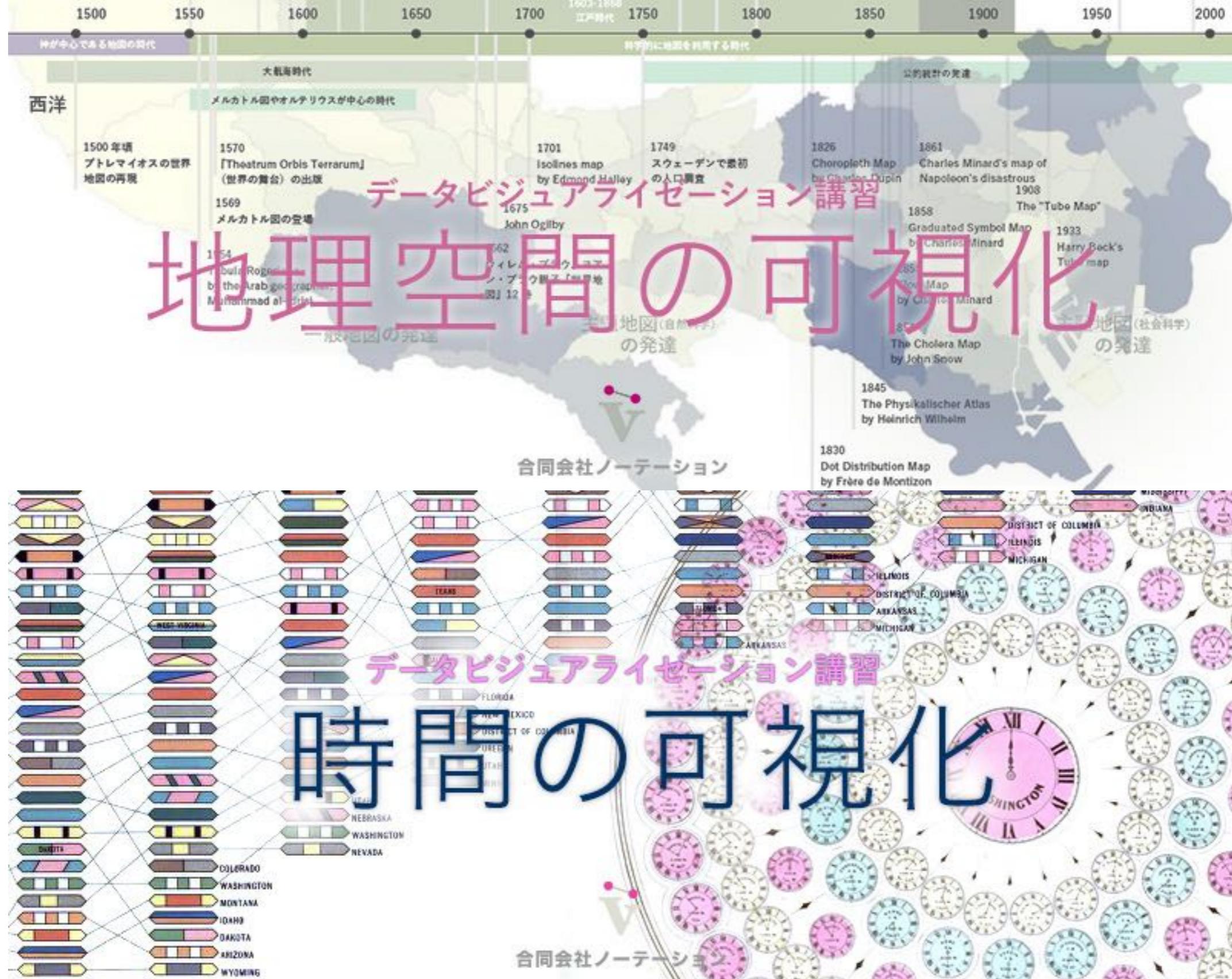
\*These figures are from the 2017 competition

# THE DATA JOURNALISM AWARDS 2018

<https://www.datajournalismawards.org/>







昨年度  
WARPデータの可視化  
ハッカソンについて

# 前段知識 (WARPとは)

## WARPとは

- ・ウェブサイトのアーカイブ
- ・2002年に始めて15年目
- ・2010年から公的機関サイトを大規模に収集



<http://warp.da.ndl.go.jp/>

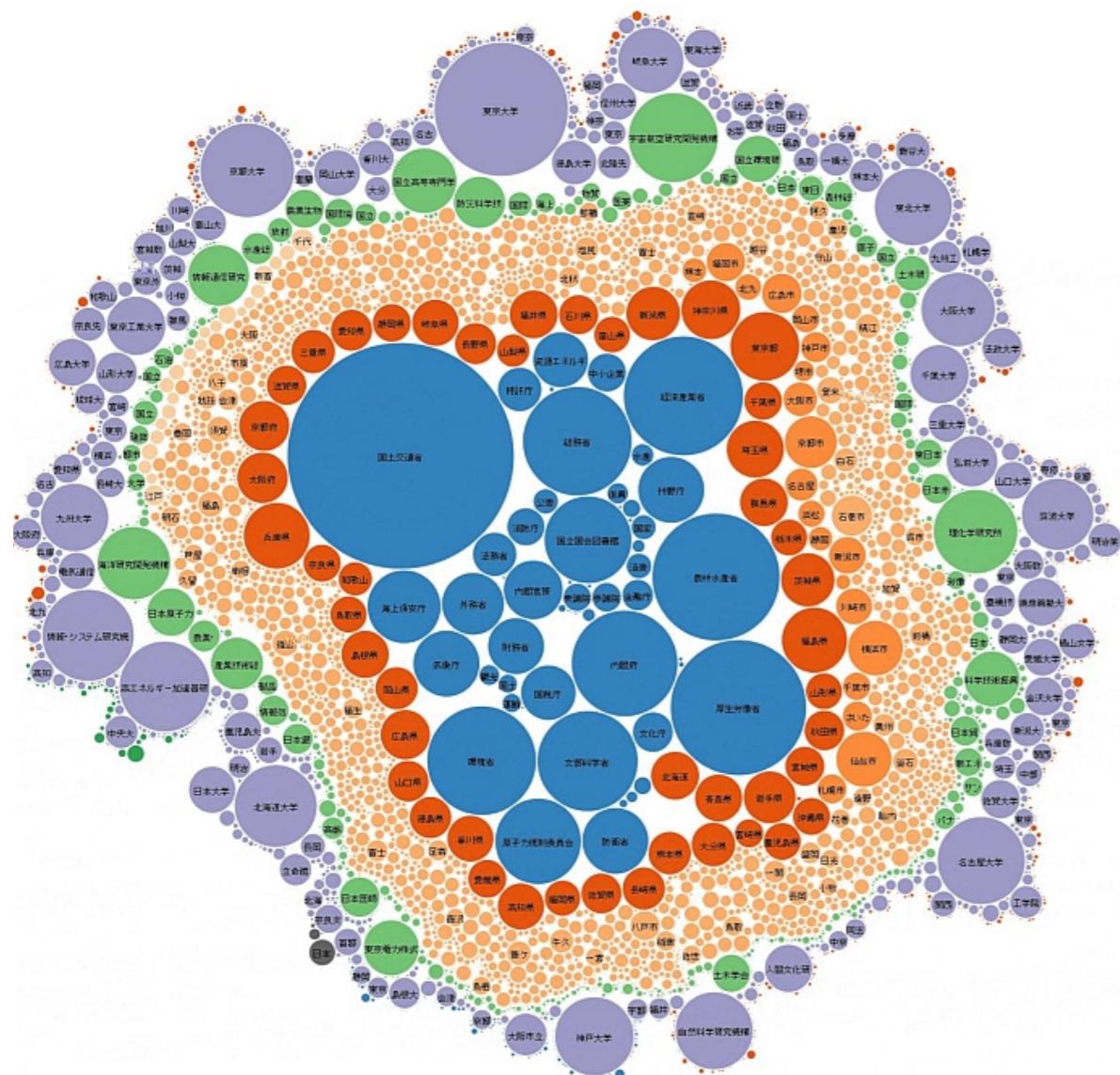
# 前段知識 (WARPとは)

## 集めているウェブサイト

区分	根拠	対象	サイト数	収集頻度
公的機関	法律	国の機関	5,400	月1回
		地方自治体		年4回
		独立行政法人		
		国公立大学		
民間	契約	公益法人、私立大学、政党、イベント、震災、電子雑誌	4,600	年1～4回

# 前段知識 (WARPとは)

容量で可視化



本日のターゲット

- 都道府県
- 政令指定都市
- 市町村
- 特別地方公共団体（東京23区を1つとして）

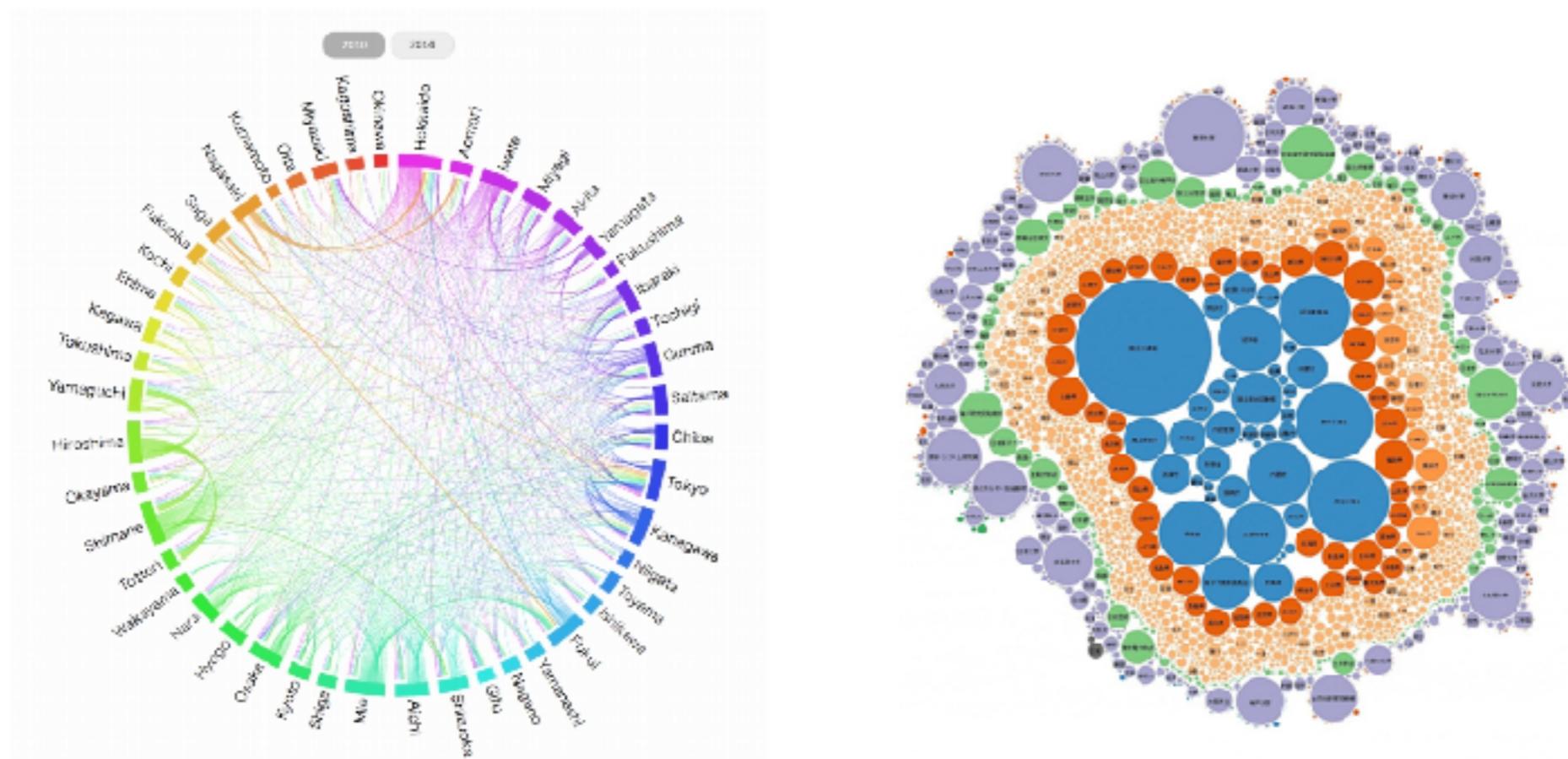
218TB/633TB

# 前段知識 (WARPとは)

## 内訳

収集対象	現存する自治体	消えた自治体	合計
都道府県	47	—	47
政令指定都市	20	—	20
市町村	1,698	989	2,687
東京23区	23	—	23
合計	1,788	989	2,777

# NDLによる可視化の事例



## 特色あるコレクション

<http://warp.da.ndl.go.jp/contents/reccommend/collection/index.html>

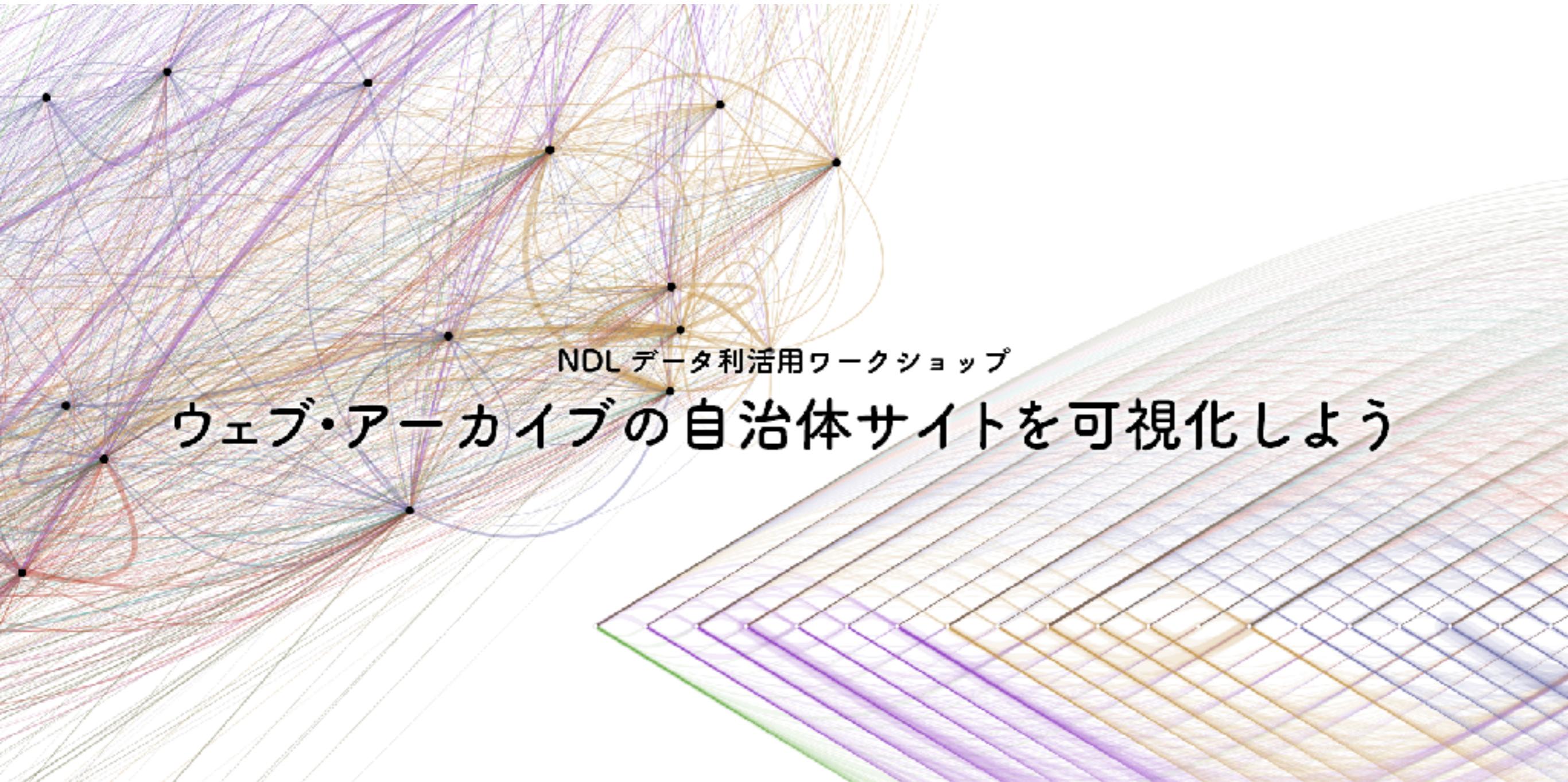
# ウェブアーカイブの意義

国立国会図書館（NDL）が取り組んでいるウェブのアーカイブ、その世界的潮流と可視化の可能性

<https://news.yahoo.co.jp/byline/yazakiyuichi/20160914-00062179/>

Web Archiving and the IIPC - Japanese

<https://vimeo.com/26878982>



## ウェブ・アーカイブの自治体サイトを可視化しよう

NDL データ利活用ワークショップ

2016年7月30日

NDLが提供するデータを可視化する「NDLデータ利活用ワークショップ～ウェブ・アーカイブの自治体サイトを可視化しよう～」を開催しました。

# 目指したこと

- ・ WARPの有効活用の事例を増やしたい
- ・ データ利活用（可視化）の事例を作りたい
- ・ 地方自治体ごとの意外な特徴や、意外な関係性をデータから探し出したい
- ・ エンジニアの実装力とデザイナーの発想力の融合

# 可視化の意義

- ・ 人間の脳は、文字よりもビジュアルな情報を処理する方が得意とされている。
- ・ 楽しい！より多くの人が関わることができる。
- ・ データ分析の結果が出てから、そのアウトリーチとしての可視化を考えるのではなく、まずは分析前に可視化してみるとところから、探索的にデータの活用の仕方や分析を考えるというアプローチ。

1日では時間が足りない…

# 参加者限定で、事前に説明会を実施



- API使い方
- アイデア出し

「NDL データ利活用ワークショップ ウェブアーカイブの自治体サイトを可視化しよう」

## アイデア例

7月27日（水）に開催したプレイベントでは、自治体サイトのデータを使った可視化について、次のようなアイデアが寄せられました！

### ■地図にマッピング系■

- 💡 サイト中の動物ワード（牛、パンダなど）の掲載ページ数のデータを使って、（xx モン GO っぽく）動物の種類や分布を地図にマッピング。[検索 API のキーワードを使用]
- 💡 「ポイ捨て禁止条例」ワードについて、掲載数の分布の経年変化を可視化することで、その政策の広がりを見せる。[検索 API のキーワード「(ポイ捨て AND 禁止)」、ファセット「都道府県」、ピボット「prefName, collectedYear」を使用]

### ■自治体の特色あぶり出し系■

- 💡 「観光」のページで「歴史」をキーワードに含むページ数から、観光資源として歴史を推している自治体を明らかにする。[検索 API のタイトル、キーワード、ファセット「都道府県名」を使用]
- 💡 「ふるさと納税」のページのお礼品を抽出して、地域の傾向を可視化する。[検索 API のタイトル、キーワード（牛肉など）、ファセット「都道府県名」、ピボット「prefName, collectedYear」を使用。とにかくいろいろなお礼品を検索してみる。]
- 💡 外国の友好都市へのリンク数を調べ、自治体の国際交流度を可視化する。[検索 API の URL「(NOT .jp)」を使用]

### ■サイトの充実度系■

- 💡 都心と地方の自治体間で、サイトのファイルサイズやコンテンツの長さに相違があるか可視化する（ウェブでも情報の東京一極集中は起こっているのか？）[メタデータと検索 API の textLength を使用]
- 💡 普段は注目されにくい「村」にフォーカスした“なんでもランキング”。例えば、サイトのファイルサイズが成長率の高さや住民1人あたりのファイルサイズの上

1日では時間が足りない…

ウェブ上にツールやデータ、  
ティップスを共有

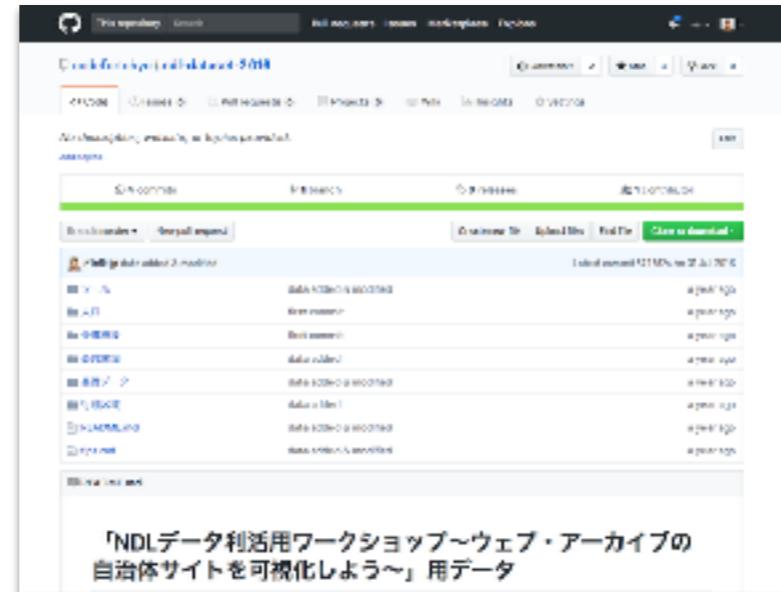
ツール



TOOL & DATA  
ツールとデータ

Code For Tokyo

データ



[https://speakerdeck.com/  
codefortokyo/tool-and-data](https://speakerdeck.com/codefortokyo/tool-and-data)

[https://github.com/  
codefortokyo/ndl-dataset-2016](https://github.com/codefortokyo/ndl-dataset-2016)

ティップス



[https://github.com/  
codefortokyo/ndl-dataset-2016/blob/master/  
tips.md](https://github.com/codefortokyo/ndl-dataset-2016/blob/master/tips.md)





カーフ・アーカイブの  
博物館サイトを  
可視化しよう！  
VISUALIZATION

カーフ・アーカイブ  
×  
博物館サイト  
×  
可視化

カーフ・アーカイブ  
×  
博物館サイト  
×  
可視化

カーフ・アーカイブの  
博物館サイトを  
可視化しよう！  
VISUALIZATION

# 成果物

- ・可視化作品とツールに大きく分けられる
  - ・可視化作品は「サイトでの発信」と「実際」の比較が多かった



# NDL Lab記事

<http://lab.ndl.go.jp/cms/?q=visualize2016>



# CodeIQ記事

<https://codeiq.jp/magazine/2016/09/45240/>

# 成果物



<https://www.facebook.com/1688505917911050/photos/a.1720600121368296.1073741828.1688505917911050/1720609354700706/?type=3&theater>

# (個人的な) 反省点

- ・ 当日の時間の短さのアピール不足。
- ・ ツールを使うとアウトプットが平準化してしまう。どこに面白さを見出すか。
- ・ 継続的な環境の提供。

当館の書誌データに基づく、  
近年の出版分野の流行り廃り

# 利用データについて

国立国会図書館サーチ

<http://iss.ndl.go.jp/>

詳細ページ

<http://iss.ndl.go.jp/books/R100000002-I027551309-00>

# 利用データについて

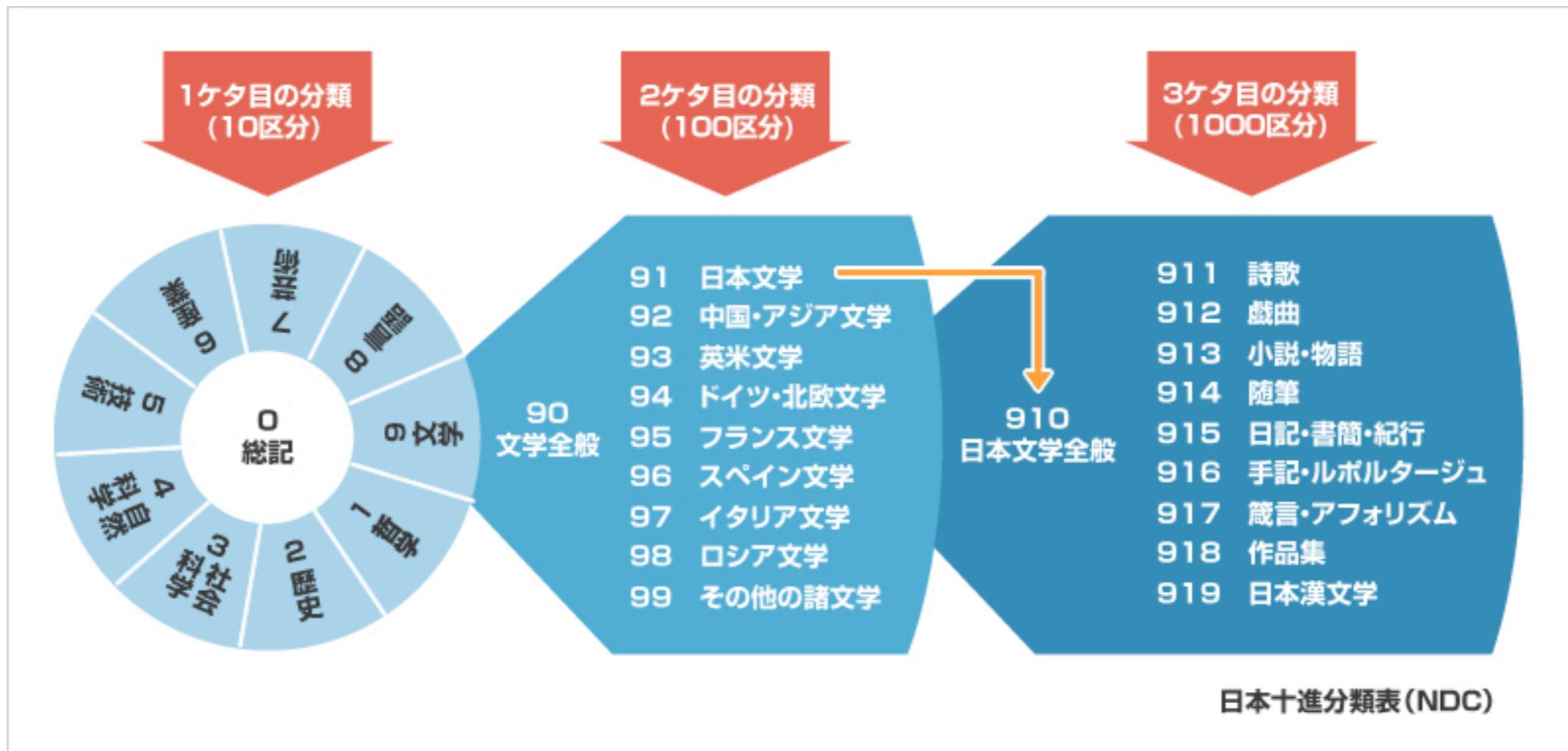
## 書誌データ

範囲：図書 material typeがBookのもののみ（逐次刊行物を含まない）／日本で出版された資料／日本語の資料／1950～2017年に出版された資料

出版流通上は逐次刊行物扱いでISSNが付与されているものを含むが、数としては多くない

項目：タイトル／著者／出版地「出版地（国名コード）」は除く／出版社／出版年／大きさ、容量等／ISBN／価格／出版年月日等／件名（キーワード）／NDLC／NDC／対象利用者／書誌ID

# 日本十進分類法 (NDC) について

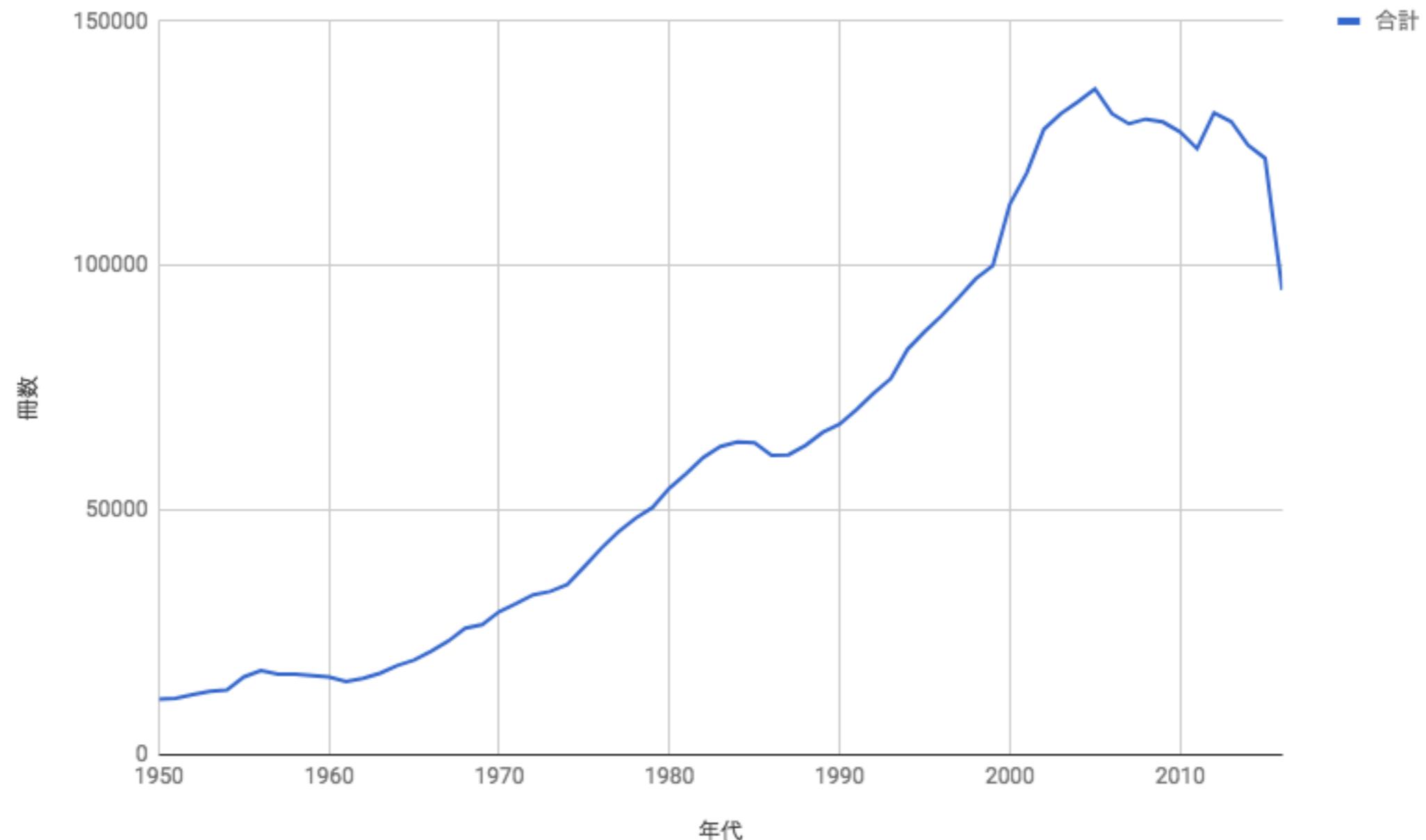


出典 <http://yamauchi-lib.jp/school/useful/>

戦後、日本で出版された全書籍の点数

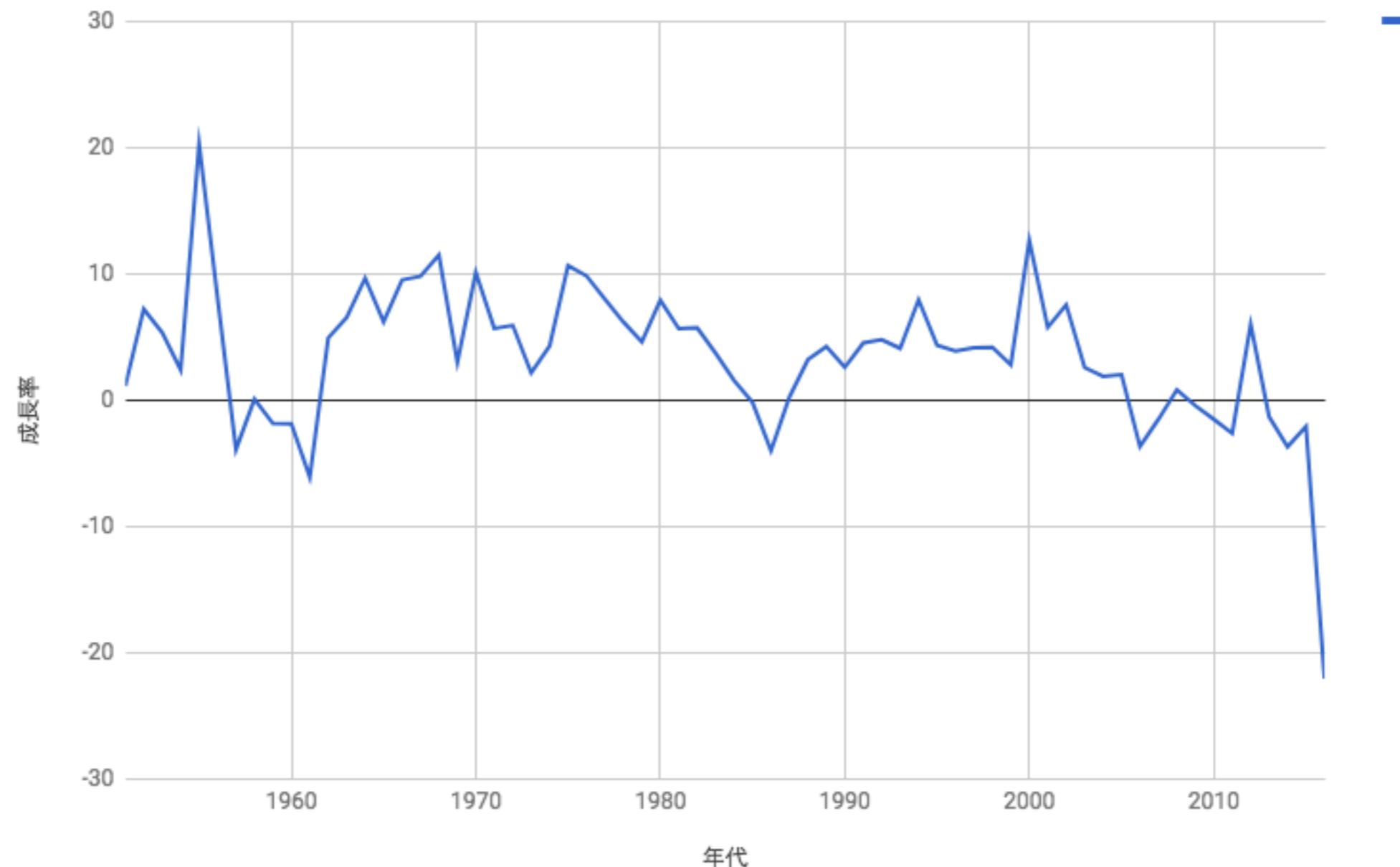
# 戦後、日本で出版された全書籍の点数

## 時系列・絶対数



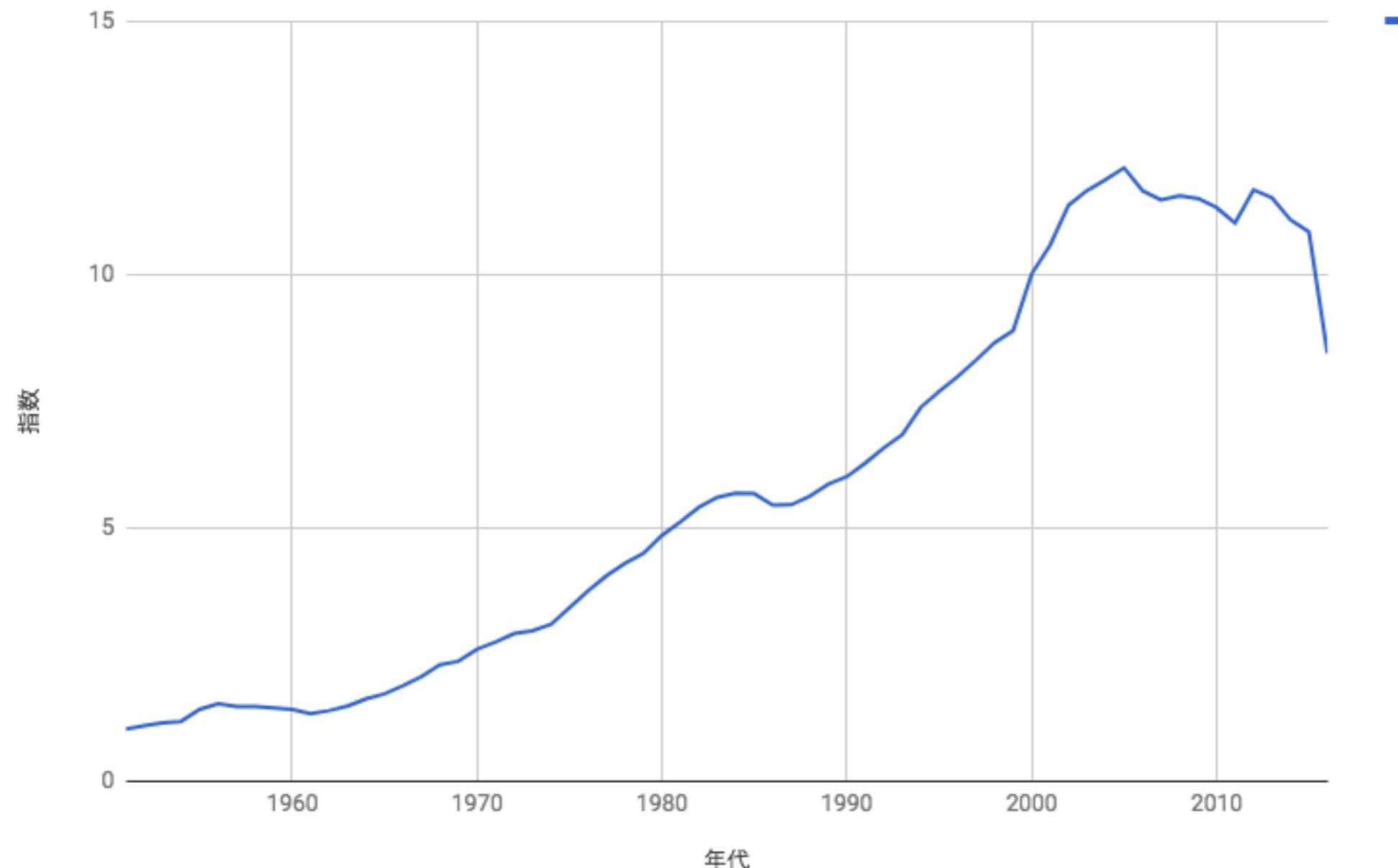
# 戦後、日本で出版された全書籍の点数

## 時系列・短期トレンド（前年比）



# 戦後、日本で出版された全書籍の点数

時系列・長期トレンド（1950年基準の指数）

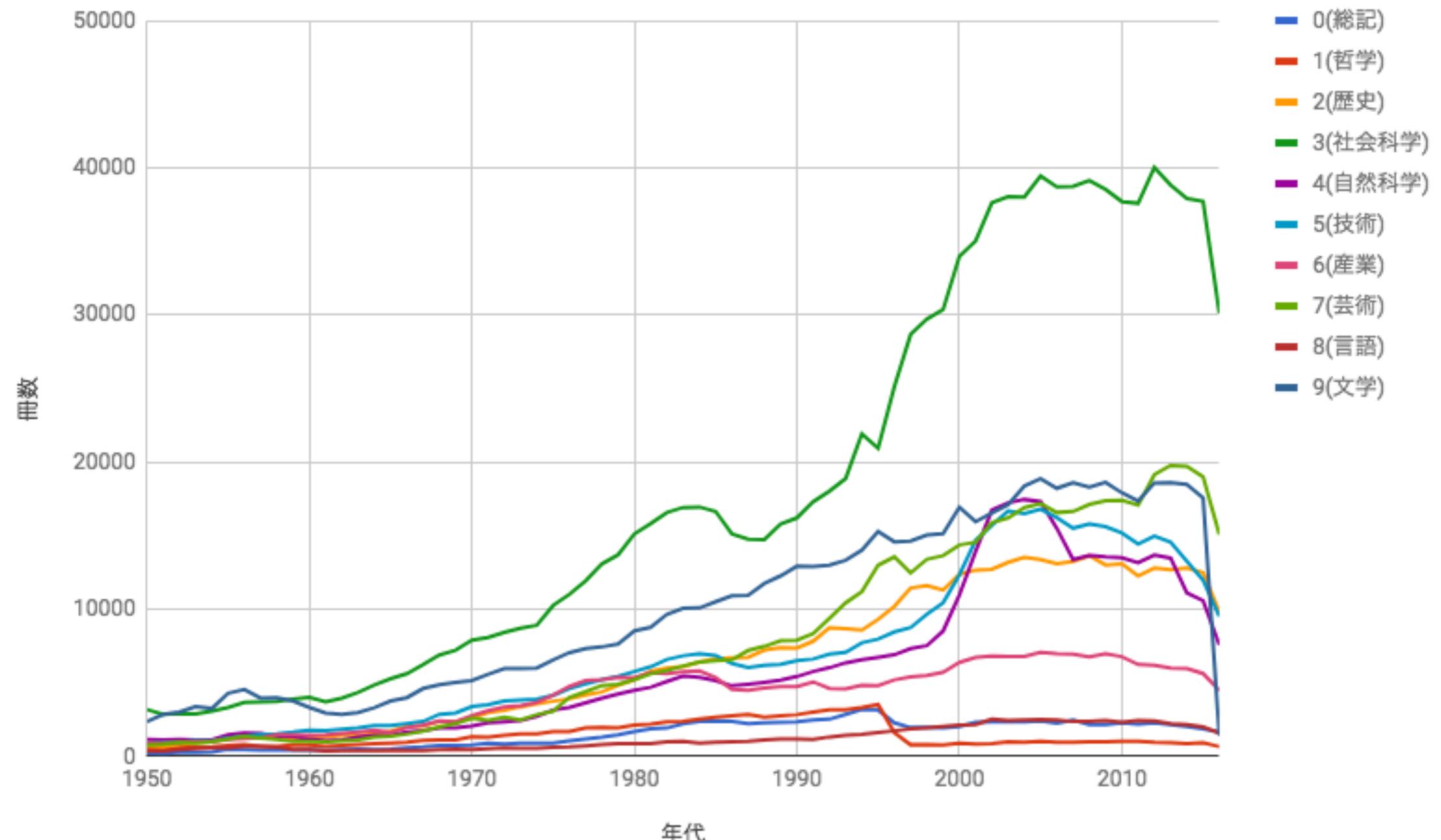


# 戦後、日本で出版された書籍の点数

時系列・NDC区分別・絶対数

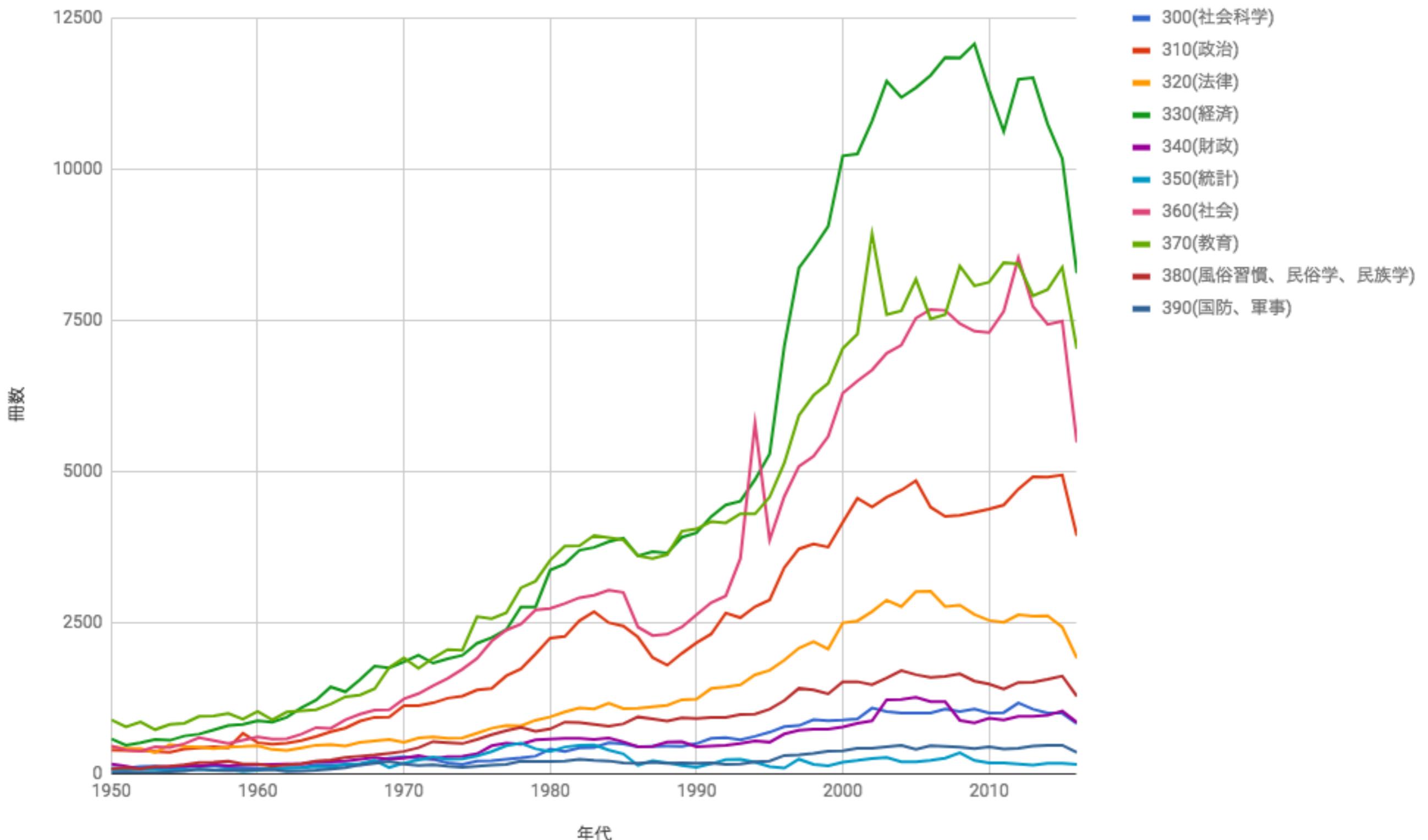
# 戦後、日本で出版された書籍の点数

## 時系列・NDC10区分・実数



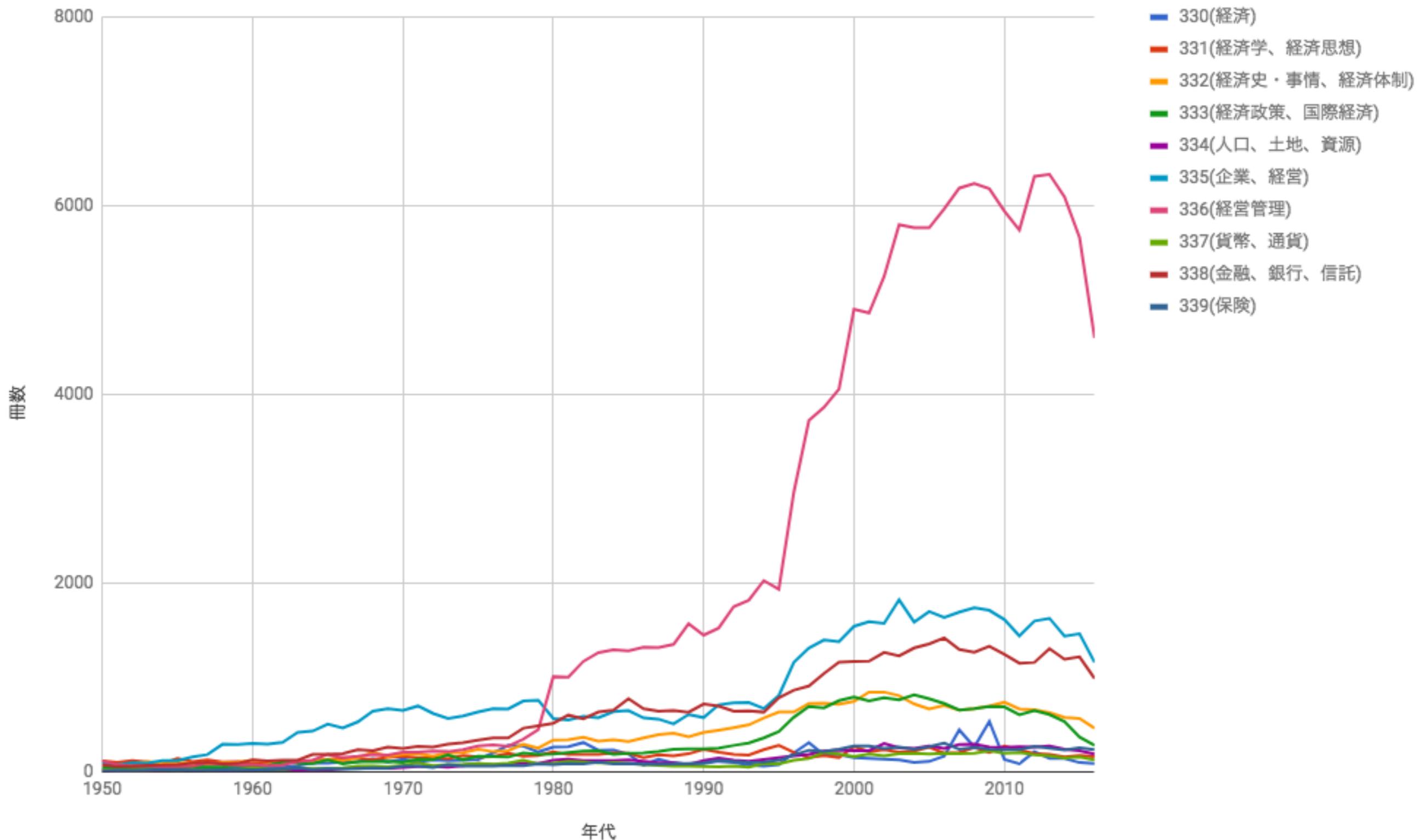
# 戦後、日本で出版された書籍の点数

時系列・NDC100区分（3.社会科学のみ）・実数



# 戦後、日本で出版された書籍の点数

時系列・NDC1000区分 (330.経済のみ) ・実数



# 「336.経営管理」に含まれる 書籍タイトル



# 336. 経営管理に含まれる書籍を 数多く出版している出版社

年間百冊以上：

講談社／小学館／中央経済社／集英社／KADOKAWA／宝島社

年間十冊以上百冊未満：

角川書店／TAC出版事業部／文芸社／PHP研究所／エンターブレイン  
／アスキー・メディアワークス／日本経済新聞出版社／秋田書店／ソ  
フトバンククリエイティブ／学研パブリッシング／日経BP社／東洋  
経済新報社税務経理協会／中経出版／ダイヤモンド社／秀和システム  
／清文社／フェアリー／スクウェア・エニックス／一迅社／三栄書房

# 多次元データの探索的可視化

作例

上場会社 一覧

<http://structure-and-representation.com/works/corporateList/>

ブラック・ブラック企業

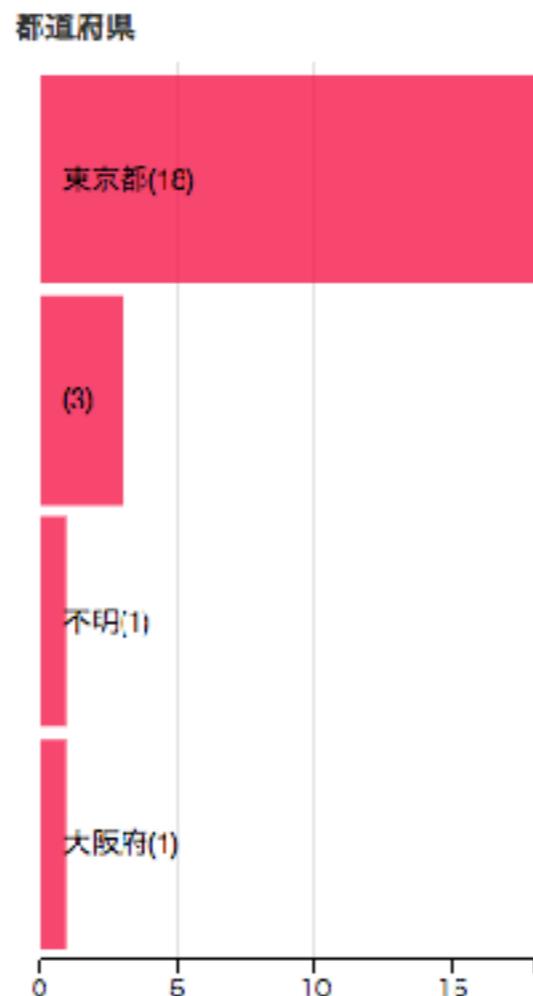
<http://structure-and-representation.com/works/blackCorporate/>

# 書誌 メタデータ ブラウザ

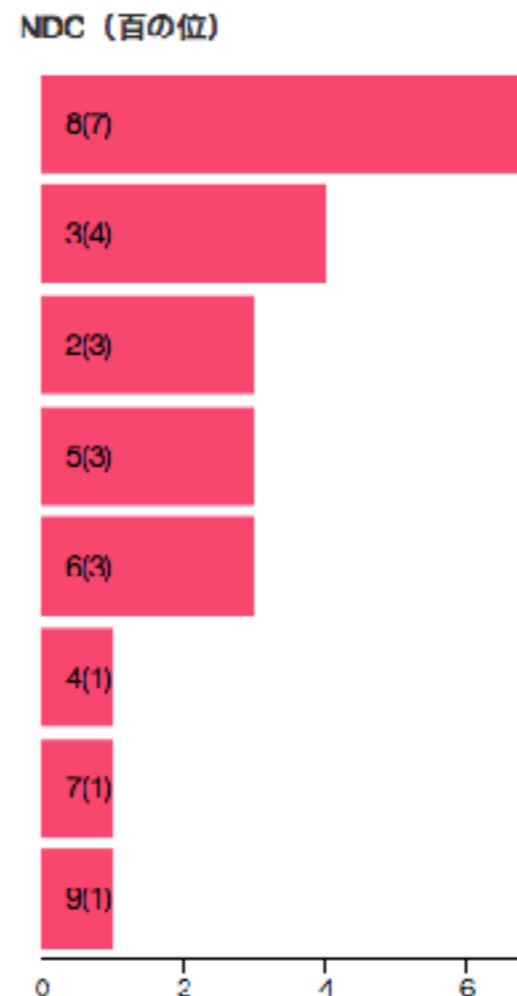
スタディ中

属性で絞り込んでください

都道府県



NDC (百の位)



出版社



## 条件に合うデータリスト

すべてのデータセットが表示されています。データセットを絞り込むには、画面右のバー・チャートから表示を残したいものをクリックしてください。

出版地\_都

道府県 タイトル

出版

年 出版社

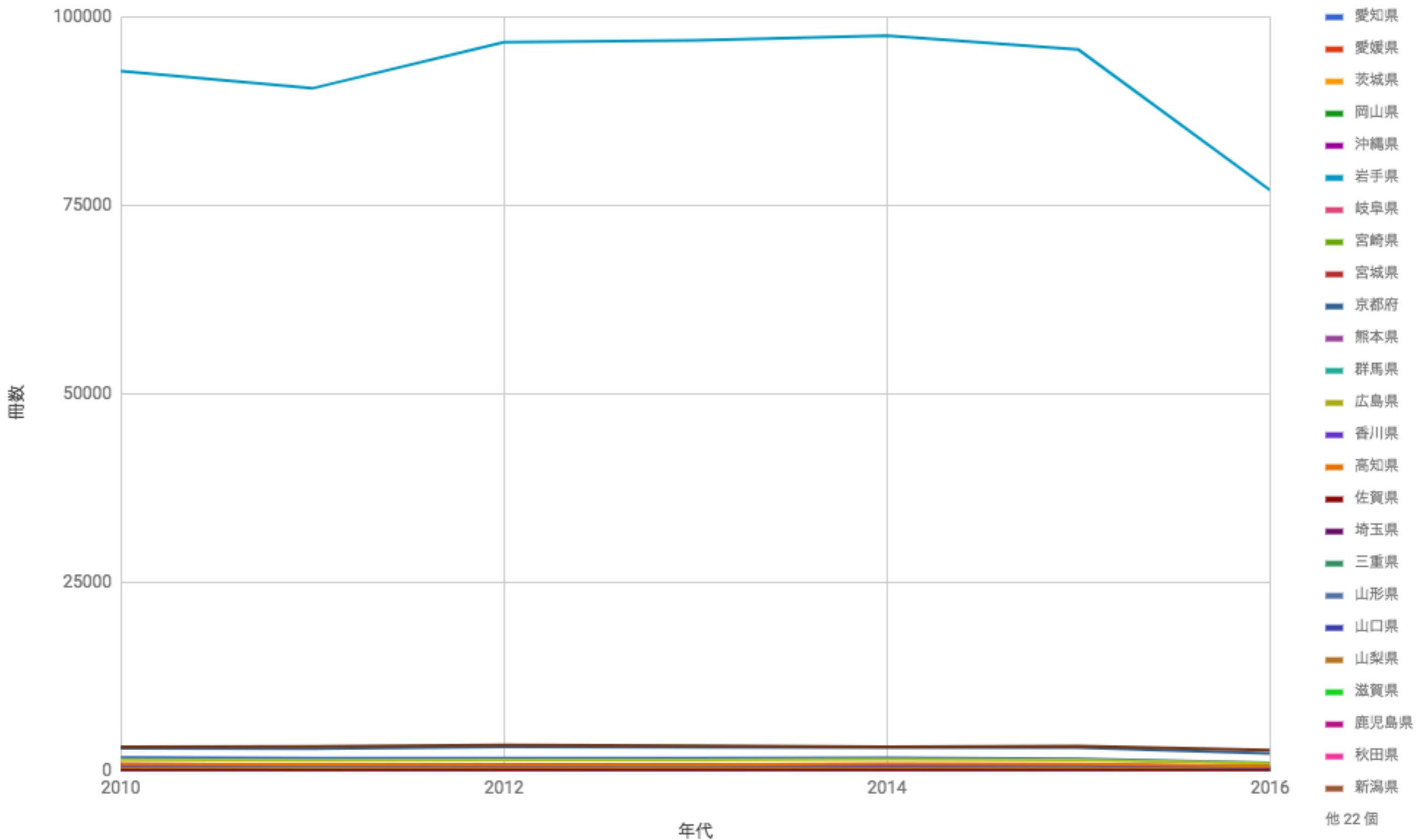
花よ・よ花よ	2017	詩画工房
甲斐・日川物語：短編小説集	2017	詩画工房
旅歌でめぐる長崎・戸町：五七五七七の心で綴る	2017	江崎厚
不明		
不明	宇宙用集積回路・高性能半導体回路の研究：静岡理工科大学教授波多野裕研究論	2017

# 戦後、日本で出版された書籍の点数

時系列・都道府県別・実数

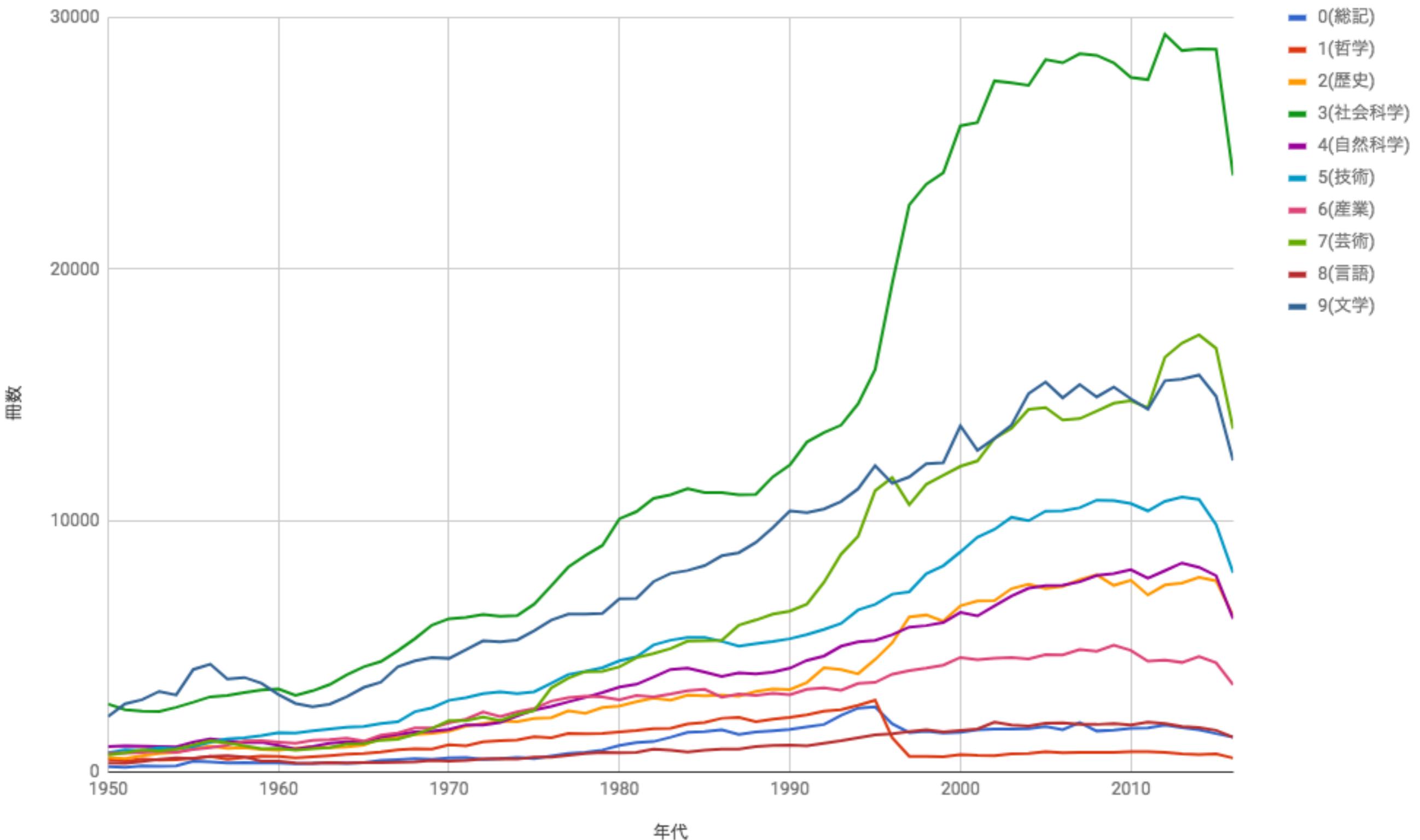
# 戦後、日本で出版された書籍の点数

## 時系列・都道府県別・実数



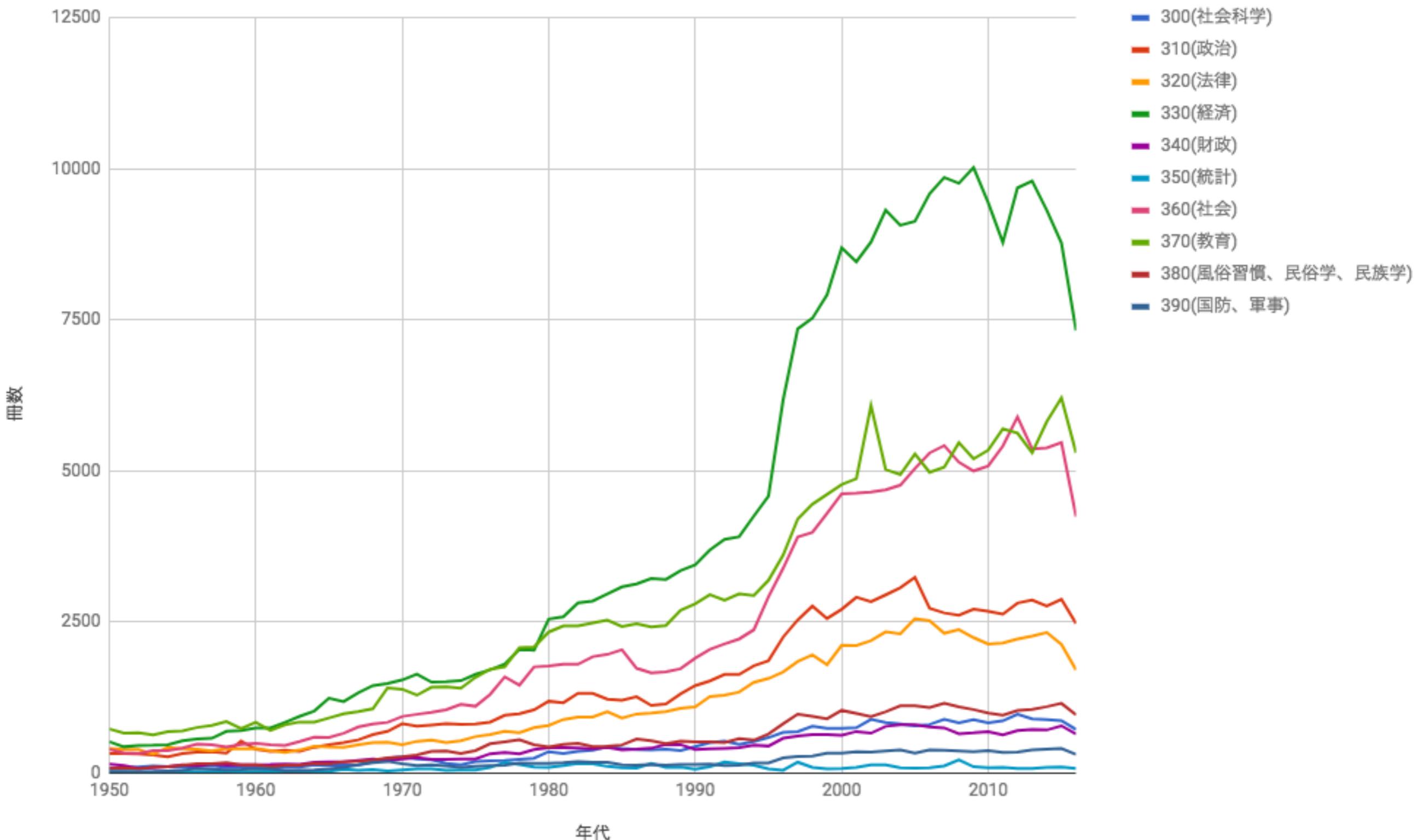
# 戦後、東京で出版された書籍の点数

## 時系列・NDC10区分・実数



# 戦後、東京で出版された書籍の点数

## 時系列・NDC100区分（3.社会科学のみ）・実数



# カートグラム（地図）による 社会科学データの表現

作例

あなたの隣の外国人

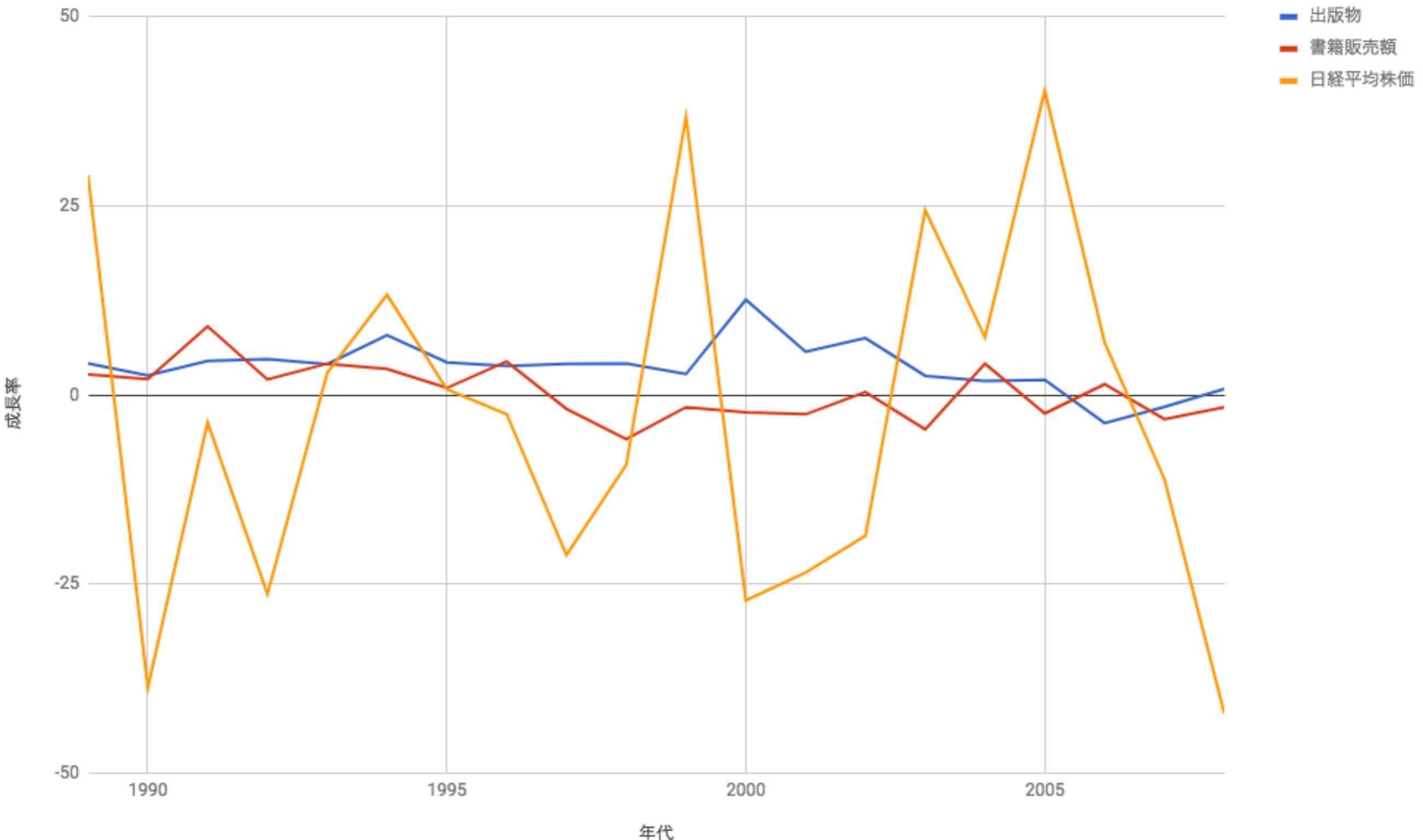
<http://tokyo.cartography.jp/foreigners/>

# NDC別 都道府県別 書誌点数 スタディ中



# 出版物点数・書籍販売額・日経平均株価

## 時系列・成長率



# このあと取り組んでみたいこと

メタ情報の充実

データドリブンな本の探し方

- ・新著と古書本の間

経済指標との比較

- ・日経平均

- ・日本経済時期

書籍の売上高との比較

- ・出版統計

# 国会での議員の発言時間

国会会議録

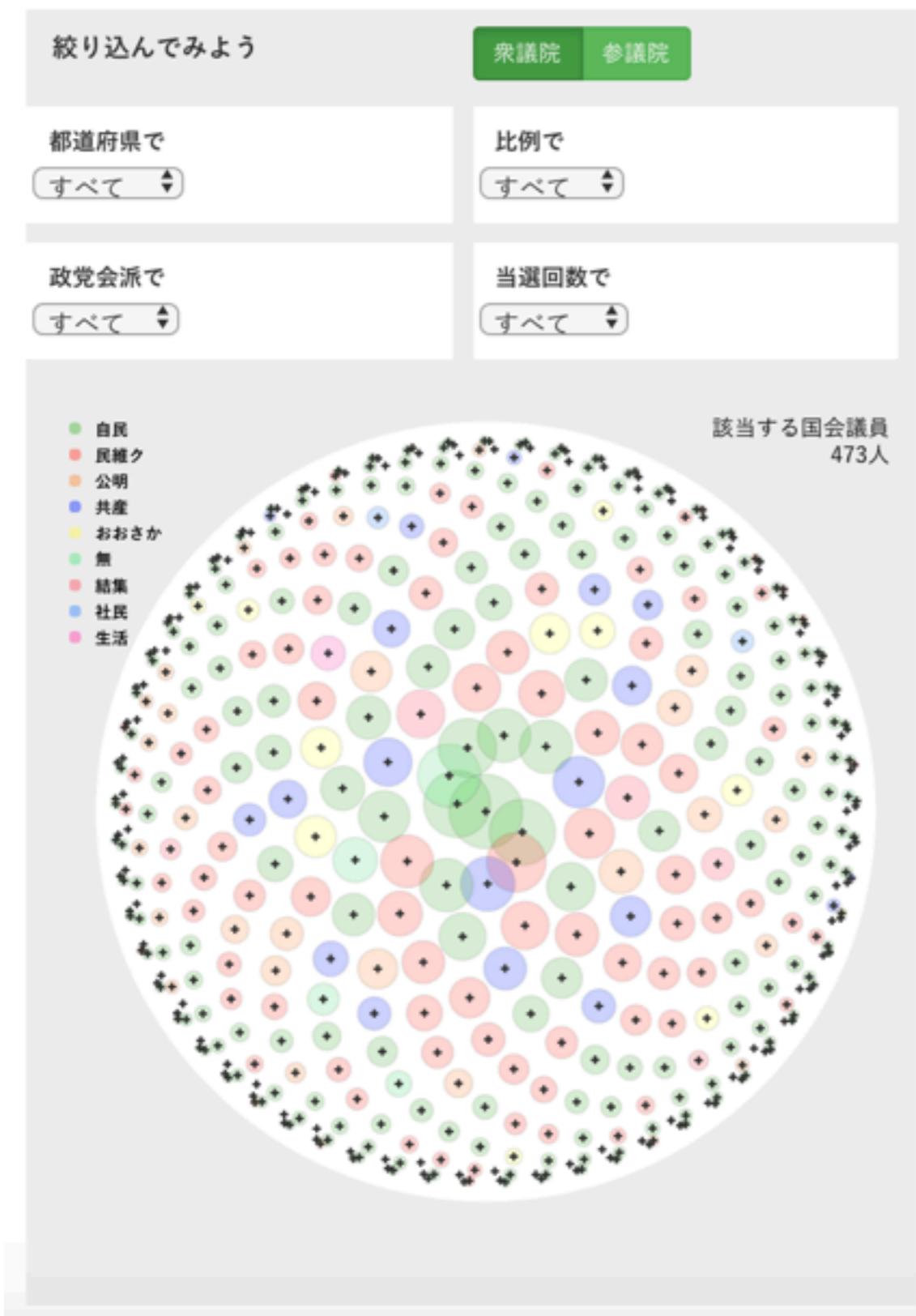
A photograph of two young boys, likely of Japanese descent, sitting at a table and looking at a tablet device. The boy on the left is wearing a dark grey t-shirt and has a white face mask hanging around his neck. The boy on the right is wearing a blue t-shirt with yellow and green accents and a graphic of two blue birds. They are both looking down at the tablet screen, which is reflecting the room. The background is slightly blurred.

最年少は中学1年生  
未来の18歳が考えた社会の課題

<https://seiji.yahoo.co.jp/youth/ideathon/tanq/>

I  
Cupid, we  
I thank the  
And I alwa

# ビジュアルでみる国會議員の発言数 国會議員は国会で何をしているの?



データ出典元：

「国会会議録」（国立国会図書館）

「衆議院議員一覧」（衆議院）

「参議院議員一覧」（参議院）

対象会期：

第189回 通常国会

平成27(2015)年01月26日～平成27(2015)年09月27日

平和安全法制各法案が審議及び可決した会期

## 1. 件名

「国会会議録発言文字数データベース」

- ・罫線は文字数に含めない。

場面の移り変わりで罫線(13文字の横棒で構成される罫線)が用いられている。

## 2. 仕様

- ・第189回 通常国会 平成27(2015)年01月26日～平成27(2015)年09月27日を対象とする。

例：\_\_\_\_\_

- ・散会を表す会議録の最終行は含めない(直接的に発言したわけない)。

例：午前十時二十六分散会

- ・発言者別に第189回 通常国会中の発言文字数を調査する。

- ・本人の発言ではない、議場と状況を表現したものは文字数に含めない。

例：(拍手)

- ・院名を指定して発言文字数を計算する。

衆議院議員は衆議院のみ。参議院議員は参議院のみとする。

一部の議員は、所属院とは他方の院で発言していることが(議題によって)あるが対象としない。

発言者とは別の議員の行動であるため、発言文字数の計算には含めない。

例：(資料提示)

## 3. 注記

- ・会議録の1行目は発言文字数に含めない。会議録の名称や開催日の情報が含まれている。

発言ではないが、発言者本人の行動であるため、そのまま残し発言文字数に含めた。

例：「189-参-政府開発援助等に関する…-1号 平成27年01月26日」

- ・半角文字・全角文字→1文字とカウント

- ・空白・改行・タブ→カウントしない

- ・括弧( )で囲まれた文字は原則、発言文字数に含めない。

議場の状況を表現したもので、発言者の直接的な発言ではない。

例：(「異議なし」と呼ぶ者あり)

(委員長退席、理事赤石清美君着席)

(国務大臣宮沢洋一君登壇、拍手)

データ取得&クレンジング：中山 圭太郎さん

<https://seiji.yahoo.co.jp/youth/ideathon/tanq/graph/>



<http://notation.co.jp/>

<http://structure-and-representation.com/>

<http://bylines.news.yahoo.co.jp/yazakiyuichi/>