



山形大学
Yamagata University

DMA101_20

データ&レポート 作成業務の効率化

山形大学 学術研究院 教授

藤原 宏司

2022.5.28

リアルタイム授業用スライド：IRプログラム第4期

第1回プレゼン発表会、お疲れ様でした！

■ What have you accomplished?

- 本プログラムにおける**最初のレポーティング**（情報提供）
- リベンジ戦は、**2022年6月18日（土）**

■ 復習：IRの役割

- 大学における諸活動のサポートに資する**情報提供を目的**に、学内外データを収集・分析（可視化）すること
- 学内外データの収集・分析の成果（結果）を伝える（**レポートを作る、資料・データを作る、発表する**）ことは、IR担当者にとって、重要な業務

なぜ、レポートティングが必要か？

- **Data-informed decision-making**
(データによって識らされた意思決定)
 - Good decision making requires good data.
 - Quality in, Quality out.
 - 意思決定者が、良い意思決定ができるような、**質の高い情報提供が必要**
- **Garbage In, Garbage Out.**
 - 質の悪いデータや、間違っている情報の提供による意思決定は、データに基づかない意思決定よりも、悪い結果をもたらしかねない。
 - データや情報の確度は、**人間のミスによる影響**「も」受ける

Reporting is...

- （日本語で書くと棘があるので、英語のまままで）

Reporting is the most **boring and important** thing in *analytics institutional research*.

- レポート作成業務は、**費用対効果が悪い**（作成側から見ると）
 - ほとんどのレポートは、一読された後にゴミ箱へ（あるある）
 - 悲しいけど、読まれないレポート「も」現実にはある
- IRが作成するレポートのほとんどが、**現状確認**
 - IRのレポートが、常に新しい発見で溢れているなんてことは無い（理想論）

レポート作成「後」あるある

■ (レポート完成後に)

送ったデータに不備がありました。

修正版を送るので、こちらのデータを反映させた新しいレポートを作ってください。

→ 心を無にして対応するしか無い。そうしないと…。

■ 今回の全学レポート、凄く良かったです！

同じようなレポートを、

学部・学科ごと、性別ごとに、過去5年分作ってください。

→ 褒められて嬉しいけど、これからの労力を考えると…。

■ もし、みなさんの次の課題が・・・？

今回のレポート、どのように作成しましたか？

- 良くあるケース（？）
 - エクセル等で分析をした結果（**図や表**）を、**パワポ※にコピー**の後、必要に応じて、**パワポ上で、図表に加筆修正**を加える
 - 上記プロセスを、レポート完成まで延々と繰り返す（Loop）
- **手作業&コピー&手作業のコラボレーション**

※ ワード等を使う方もいると思いますが、今回は、パワポと想定します。

分析作業に用いたエクセルファイルは？

- オリジナルのエクセルファイルに、追加された関数や、作成された図表が散りばめられた「本人以外は分からない」状態になっていることが多い？

- 本人でさえも、3ヶ月も経てば「何をしたか分からなくなっている」のでは？

さらに

- エクセルで図表作成 → パワポにコピペ → パワポ上で修正

なんてことが起きていると、
同じレポート、同じ図表が再現できないこともある

→ 考えられる理由：作業手順が記録されていない

Creating an Efficient Reporting Workflow

■ IR※が作成するレポートは、

- エラーが無く（人が作るのでミスは起きるけど）
- 全く同じデータや図表を「第3者が再現可能」で
- 再利用も可能

な手順（ワークフロー）で作成されることが望ましい

※ IRだけではなく、他の部署が作るレポートにも当てはまると思います。

業務を継続していくために

- レポート完成までの思考、行ったトライ&エラー、実際の作成手順等の「**全て**」を文章化
 - 特に「人事異動制度」を運用している組織では、これが重要
 - これができていない引き継ぎは、**質の悪い伝言ゲーム**
- ちなみに、米国の大学では、政府等に提出するレポートに使用した「**元データ**」や「**KPI算出手順**」を「**再利用可能な形**」で保存している
 - いつでも、政府に提出した「**データ、KPIが再現**」できる
- 米国の大学（例：MnSCU）における**DX化の基盤**は、1990年代後半から2000年代前半にかけて整備
 - その時のトップによる意思決定等のお話は、今後の授業で

再現 & 再利用：なぜ重要？（例）

■ レポートの再現

- **検証、比較**のために必要
- 例：過去のレポートにあったKPIを、再現できますか？
（集計値として保存していませんか？ → ダメな方法）

■ レポート（作成フロー）の再利用

- 修正データへの対応
- 同じ形式のレポートを作成（再生産）
 - 全学用、学部用、学科用、性別ごと → **手順は全部同じ**

■ 再現&再利用できるレポートを作るには？

- 手作業からの脱却
- **コピー汚染**を起こさない（← ネットから拾った表現、イイね！）

(再掲) 比較 : プログラム (R) vs. 手作業

比較項目	R	手作業
単純な手作業	自動化できる	ひたすらコピペ
エラー	規則的	ランダム
エラーの特定	比較的容易	難しい (作業をやり直す?)
同じ作業の再現 (追加リクエスト等)	簡単	根性と時間があれば (人間関係が心配)
スキルの共有	コードを共有する	手作業のコツは 伝えられるけど...
ダブルチェック	同僚にコードを 見てもらう	同僚に同じ作業を してもらう?
習熟スピード	時間がかかる	それほどかからない
根性	Rを覚えるのに必要	作業をするのに必要
充実を感じる時	プログラムが動いた!	徒労感の方が 大きいのでは?

R Markdownによるレポーティング



注意点（重要）

- 本プログラムの標準学修時間は、**60時間**と少なく、他にも学修すべき科目群もあることから、これから紹介する内容の「全て」を、プログラム修了（9月末）までにカバーできるかは未定です。
 - **R Markdownの基礎（HTMLレポートの作成）を扱います。**
- **体調**や時間が許せば、10月～翌年2月までの間に、**不定期**で「**R Tips & Tricks**」というミニ講義シリーズを開講します。そこで、扱える「かも」しれません。
 - 担当：藤原&田中先生
 - 受講は強制ではありません。

デモンストレーション #1

- みなさんが**実際に分析したデータ**を使って、**C学部の「授業外学習時間に関する分析」レポートをMS Wordに出力**
 - RStudio → レポート出力 → MS Word
- **レポートの内容**
 - フロントページ
 - C学部における**アンケート回答率の表**（学科、性別ごと）
 - 回答率 < 0.9 を赤字表示
 - **授業外学習時間に関するヒストグラム**（性別ごと）
 - **大学全体の平均授業外学習時間およびC学部における性別ごとの平均授業外学習時間も描画**
などなど

MS Word 出力例 (1)



- 2020 年度 IR 大学入学生アンケートについて
- C 学部：授業外学習時間に関する分析
- 藤原宏司（山形大学）
- 2022-05-27 05:51:36
- エグゼクティブサマリー
 - C 学部に関するレポートのハイライト（3つくらいが理想）を書きましょう。
 - 客観的な事実を簡潔に箇条書きで書きましょう。
 - 詳しくは今後の授業内で説明します。
- MS Word によるレポートングについて
 - 結論から言うと、HTML で作成するよりも制限が多く、WEB 上における資料も少ない。
 - カラー文字やハイライトが使えないのが痛い。
 - 開発者の方も「まずは HTML レポートを検討した方が良い」と仰っていることから、授業では「HTML によりレポートング」先に勉強します。



アンケート回答率（全てのアンケートが対象）

Department	Gender	入学者数	回答者数	回答率
1	Female	18	16	0.89
1	Male	53	51	0.96
1	学科ごと	71	67	0.94
2	Female	17	17	1.00
2	Male	62	60	0.97
2	学科ごと	79	77	0.97
3	Female	9	8	0.89
3	Male	45	43	0.96
3	学科ごと	54	51	0.94
4	Female	24	20	0.83
4	Male	41	38	0.93
4	学科ごと	65	58	0.89
5	Female	21	20	0.95
5	Male	23	19	0.83
5	学科ごと	44	39	0.89
学部全体	Female	89	81	0.91
学部全体	Male	224	211	0.94
学部全体	学科ごと	313	292	0.93

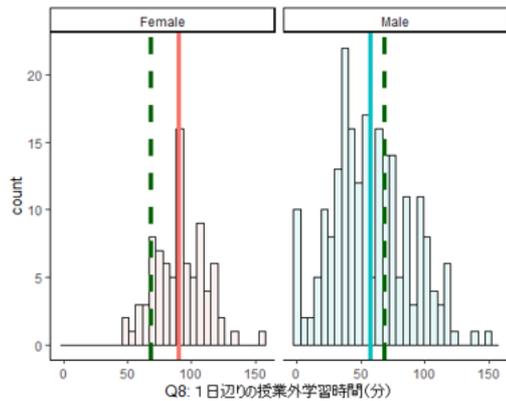
Note: 回答率が「0.9 未満」を赤字で表示している。

実際のWordファイルは、Teamsで共有します

MS Word 出力例 (2)



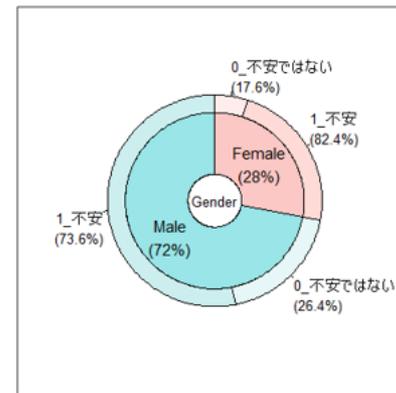
• c 学部: 1日辺りの授業外学習時間(分) ←



Note: 破線は大学全体の平均授業外学習時間、実線は、その学部における性別ごとの平均授業外学習時間である。 ←



• c 学部: 性別 vs. 不安感 (Q9) ←

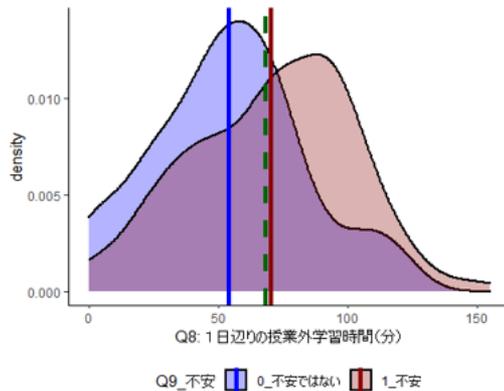


Note: Q9 における回答のうち、「1=とても感じている」および「2=感じている」を「1_不安」にまとめ、「3=あまり感じていない」および「4=全く感じていない」を「0_不安ではない」とした。 ←

MS Word 出力例 (3)



・C 学部：不安感 (Q9) vs. 1 日辺りの授業外学習時間 (分) の分布



Note: 破線は大学全体の平均授業外学習時間、実線は、その学部における不安グループごとの平均授業外学習時間である。



・資料：1日辺りの授業外学習時間 (分)

College	Gender	入学者数	平均値	標準偏差	最小値	中央値	最大値
A	Female	176	74.5	10.1	40	75	95
A	Male	187	49.6	20.3	0	50	100
A	学部全体	363	61.6	20.4	0	65	100
B	Female	129	77.3	15.4	35	80	115
B	Male	65	60.2	14.2	15	60	90
B	学部全体	194	71.5	17.0	15	70	115
C	Female	89	90.1	19.5	50	90	155
C	Male	224	57.6	30.8	0	55	150
C	学部全体	313	66.7	31.6	0	70	155
D	Female	103	84.0	9.7	65	85	115
D	Male	405	64.6	24.1	0	65	150
D	学部全体	508	68.5	23.3	0	70	150
E	Female	95	90.7	11.8	60	90	120
E	Male	77	70.4	16.7	30	70	100
E	学部全体	172	81.6	17.4	30	85	120
大学全体	Female	592	81.7	14.8	35	80	155
大学全体	Male	958	60.2	24.9	0	60	150
大学全体	学部全体	1,550	68.4	24.0	0	70	155

→ デモンストレーション (コピペをしていない点に注目してください)

レポート作成「後」あるあるへの対応（1）

■ データにエラーがありました編

- 男性の授業外学習時間：60分以上 → 0分でした
- 男性で「Q9 = “1とても感じている”」という回答が、「Q9 = “4全く感じていない”」でした

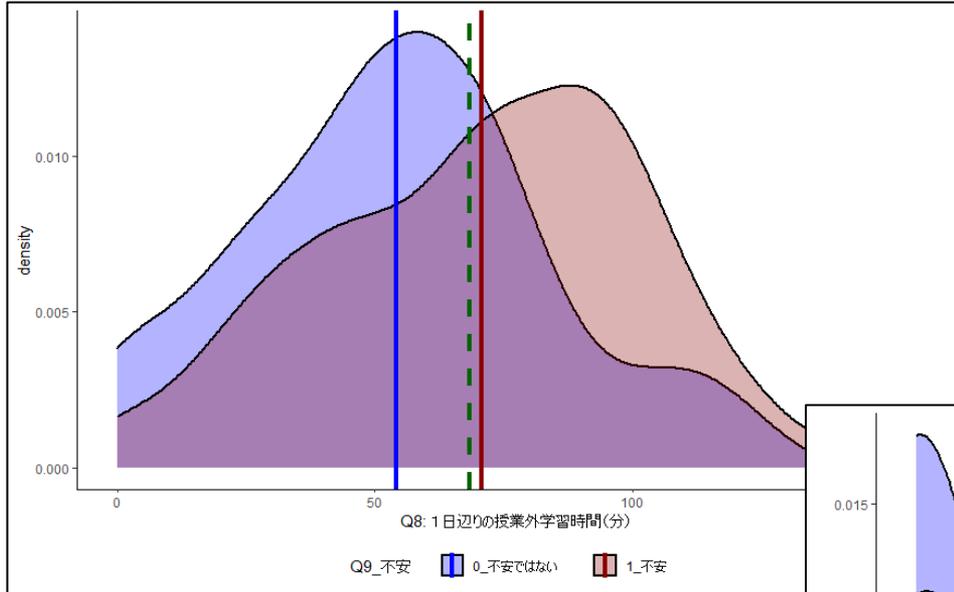
■ 対応：「分析用データファイル」を差し替えるだけ

- 変更箇所：**1行**
- 作業時間：**数秒**

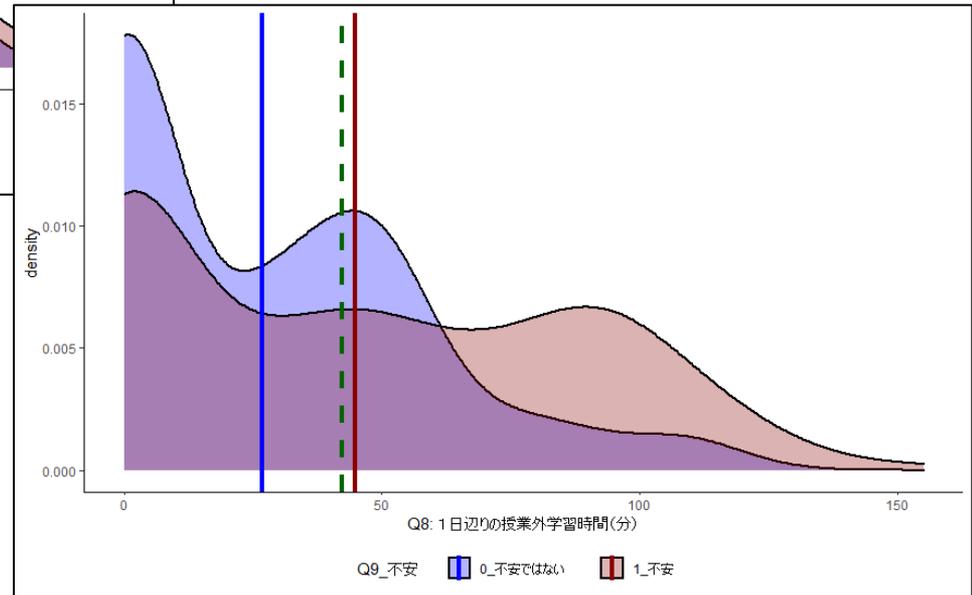
→ 全ての図表が、変更されたデータをもとに、作り直されます

グラフの変更例

変更前



変更後



デモ後に作成されたWordファイルは、Teamsで共有します

レポート作成「後」あるあるへの対応（2）

■ 下記の表、情報（列）が不足してませんか？

1日辺りの授業外学習時間(分)

College	Gender	入学者数	平均値	標準偏差	最小値	中央値	最大値
A	Female	176	74.5	10.1	40	75	95
A	Male	187	49.6	20.3	0	50	100
A	学部全体	363	61.6	20.4	0	65	100
B	Female	129	77.3	15.4	35	80	115
B	Male	65	60.2	14.2	15	60	90
B	学部全体	194	71.5	17.0	15	70	115
C	Female	89	90.1	19.5	50	90	155
C	Male	224	57.6	30.8	0	55	150
C	学部全体	313	66.7	31.6	0	70	155
D	Female	103	84.0	9.7	65	85	115
D	Male	405	64.6	24.1	0	65	150
D	学部全体	508	68.5	23.3	0	70	150
E	Female	95	90.7	11.8	60	90	120
E	Male	77	70.4	16.7	30	70	100
E	学部全体	172	81.6	17.4	30	85	120
大学全体	Female	592	81.7	14.8	35	80	155
大学全体	Male	958	60.2	24.9	0	60	150
大学全体	学部全体	1,550	68.4	24.0	0	70	155

デモ後に作成されたWordファイルは、Teamsで共有します

レポート作成「後」あるあるへの対応（3）

- **他の学部用のレポートも作成してください編**
- **対応①：「分析用のデータ」指定を変えるだけ**
 - **変更箇所：1行**
 - **作業時間：数秒**

→ **全ての図表が、指定された学部のデータをもとに、作り直されます**
- **上記作業を、学部分だけ行います**
- **対応②：ループで一括処理**
 - **授業では扱いません → 「R Tips & Tricks」？（未定）**

デモ後に作成されたWordファイルですが、共有しません。
理由：実際にデータ分析をするのも練習です！

R Markdown × HTMLレポートティング



デモンストレーション #2

- 学生1人につき1ファイルの成績レポートを**自動作成**します
 - 学生数 = 100人 → 100個のファイルを自動作成
 - 僕のノートPCだと、6分程度かかるはず
 - その間に**休憩**しましょう！
 - レポートの内容は、その後に説明します
 - ついでに、**データ可視化に関するお話**も少しだけ

例：100ファイル作成までにかかった時間

名前 ▲	サイズ	項目の種類 ^	更新日時
出力_主要5教科成績比較_001.html	952 KB	Chrome HTML D...	2022/05/28 11:52:50
出力_主要5教科成績比較_002.html	953 KB	Chrome HTML D...	2022/05/28 11:52:53
出力_主要5教科成績比較_003.html	932 KB	Chrome HTML D...	2022/05/28 11:52:56
出力_主要5教科成績比較_004.html	932 KB	Chrome HTML D...	2022/05/28 11:52:58
出力_主要5教科成績比較_005.html	952 KB	Chrome HTML D...	2022/05/28 11:53:01
出力_主要5教科成績比較_006.html	953 KB	Chrome HTML D...	2022/05/28 11:53:03
出力_主要5教科成績比較_007.html	952 KB	Chrome HTML D...	2022/05/28 11:53:06
出力_主要5教科成績比較_008.html	931 KB	Chrome HTML D...	2022/05/28 11:53:09
出力_主要5教科成績比較_009.html	951 KB	Chrome HTML D...	2022/05/28 11:53:11
出力_主要5教科成績比較_010.html	932 KB	Chrome HTML D...	2022/05/28 11:53:14

出力_主要5教科成績比較_090.html	952 KB	Chrome HTML D...	2022/05/28 11:56:43
出力_主要5教科成績比較_091.html	952 KB	Chrome HTML D...	2022/05/28 11:56:46
出力_主要5教科成績比較_092.html	951 KB	Chrome HTML D...	2022/05/28 11:56:48
出力_主要5教科成績比較_093.html	931 KB	Chrome HTML D...	2022/05/28 11:56:51
出力_主要5教科成績比較_094.html	951 KB	Chrome HTML D...	2022/05/28 11:56:54
出力_主要5教科成績比較_095.html	952 KB	Chrome HTML D...	2022/05/28 11:56:56
出力_主要5教科成績比較_096.html	951 KB	Chrome HTML D...	2022/05/28 11:56:59
出力_主要5教科成績比較_097.html	931 KB	Chrome HTML D...	2022/05/28 11:57:01
出力_主要5教科成績比較_098.html	932 KB	Chrome HTML D...	2022/05/28 11:57:04
出力_主要5教科成績比較_099.html	931 KB	Chrome HTML D...	2022/05/28 11:57:07
出力_主要5教科成績比較_100.html	953 KB	Chrome HTML D...	2022/05/28 11:57:09

HTMLレポートの例（1）



1. シナリオ
2. Radar Chart
3. Lollipop Chart
4. Side-by-Side

主要5教科アセスメント: Test 1 vs. Test 2

学生番号: 001

藤原宏司（山形大学）

Last Updated: 2022-05-28

1. シナリオ

- 下記の2つのグラフ（Radar ChartとLollipop Chart）は、**同じデータを、それぞれ異なる手法で可視化**したものです。主要5教科における2回のテストスコアをグラフ化したものになりますが、どちらのグラフの方が**クライアントにとって見やすいか、理解しやすいか**、考えてみてください。
- このデモンストレーションでは、学生1人つき1ファイルの成績レポートを自動作成します。テストデータは100人分ありますので、100つのファイルが短時間で出来上がります。

2. Radar Chart

学生番号: 001

- 定番?的な比較方法



HTMLレポートの例 (2)

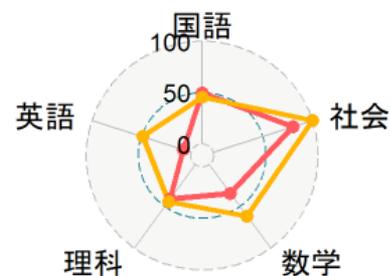
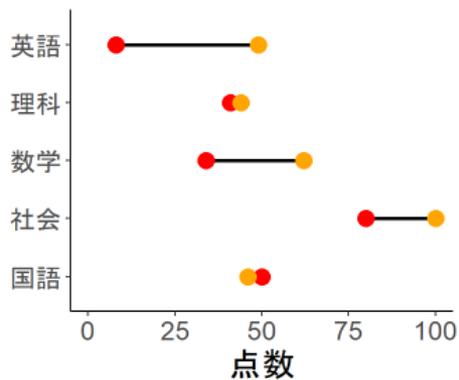


1. シナリオ
2. Radar Chart
3. Lollipop Chart
4. Side-by-Side

4. Side-by-Side

学生番号: 001

- 並べてみました。比較しやすい？



Test ● Test_1 ● Test_2 —●— Test_1 —●— Test_2

- **本プログラムでは、HTMLレポートの作成方法をメインに学習します。**

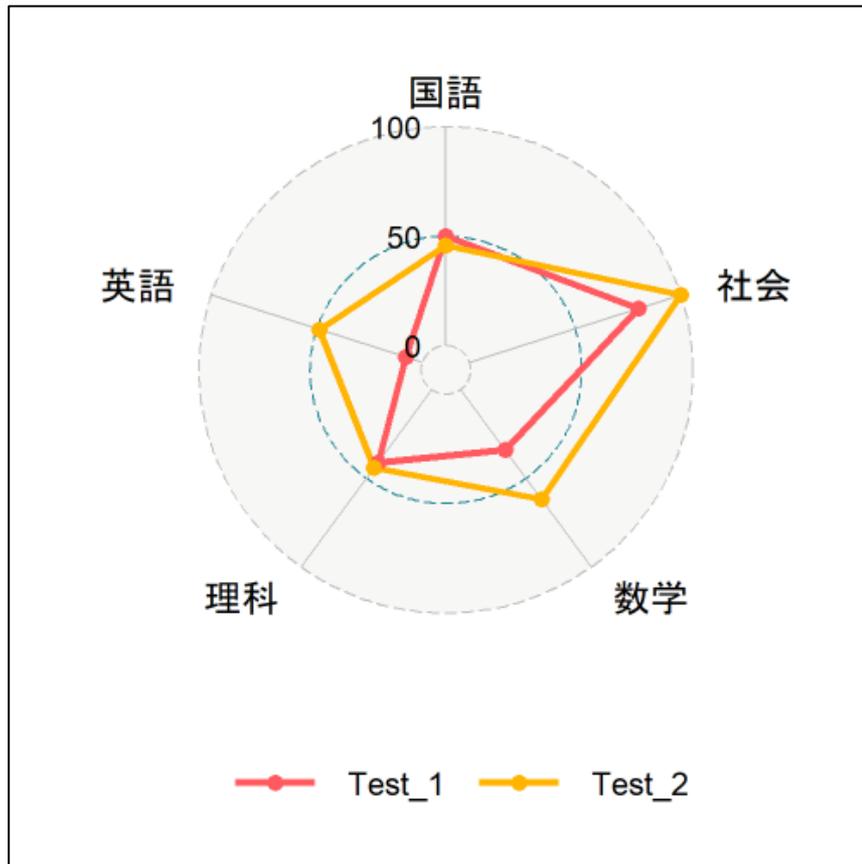
データの可視化方法

Radar Chart vs. Lollipop Chart



データの可視化 : Radar Chart

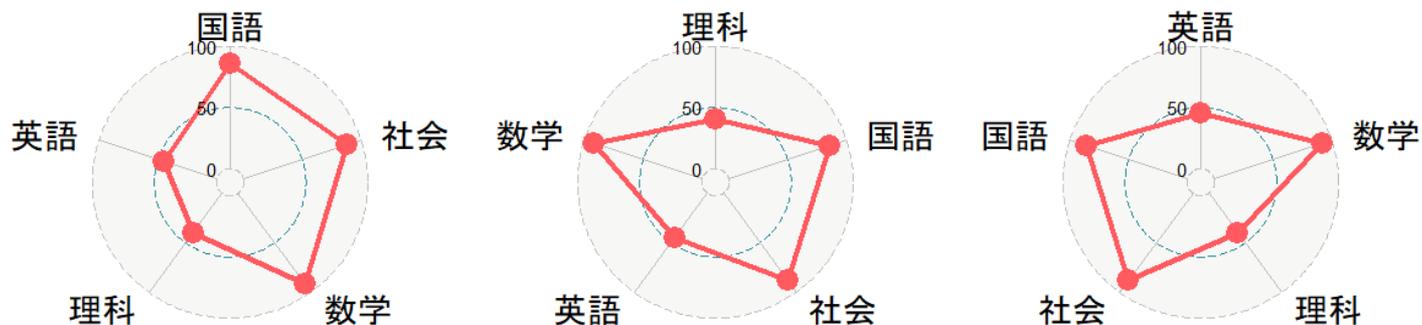
- 成績等、個人が持つ**複数のパラメーター**を描画する時に、良く使われるグラフ形式



- 定番的なグラフですが、**批判的な意見**も多いです
- **どこに問題があるか、考えてみましょう**

Radar Chartの代表的な問題点

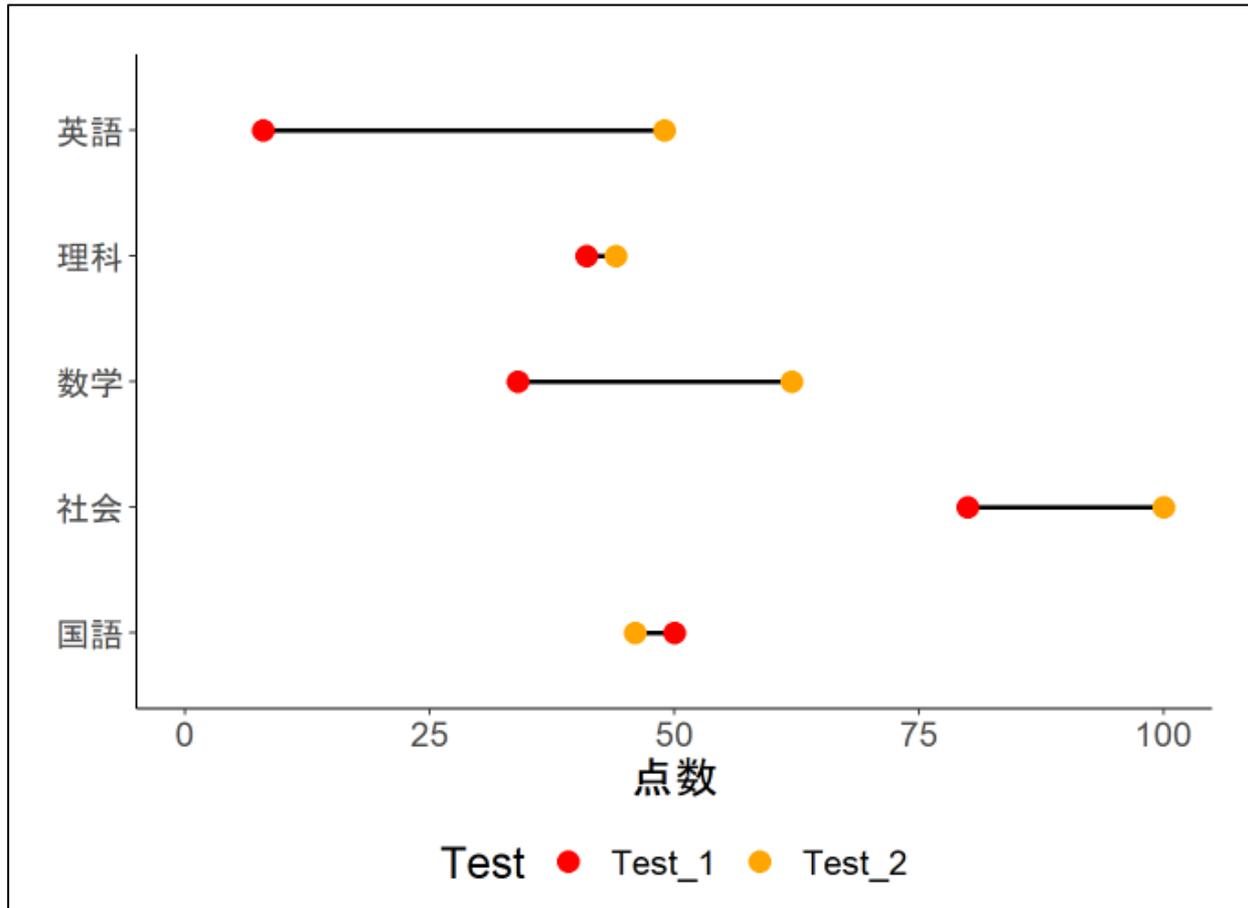
- 以下の3つのグラフは、**全て同じデータから描画**されています。



- **グラフから受ける印象は同じですか？**

データの可視化 : Lollipop Chart

- このような可視化方法はどうですか？（p.24と同じデータ）



BIレポート



BI (Business Intelligence) レポート

- 本プログラムでは、**Microsoft Power BI**を使ったレポートの作成について、学習します。
 - 使用データ：今回のプレゼン発表回用データ
 - 作成後のイメージは次のページ
 - URLは、Teamsで共有します

Power BIレポートの例 (1)

IR大学「オンライン授業に関するアンケート調査」

対象：2020年度入学生 (1,550名)

学部

- すべて選択
- A
- B
- C
- D
- E

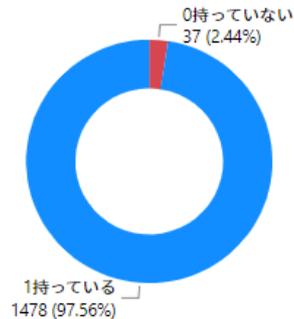
学科

- すべて選択
- 1
- 2
- 3
- 4
- 5

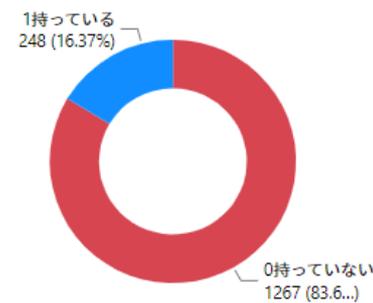
性別

- すべて選択
- Female
- Male

Q1. スマートフォンの所持状況

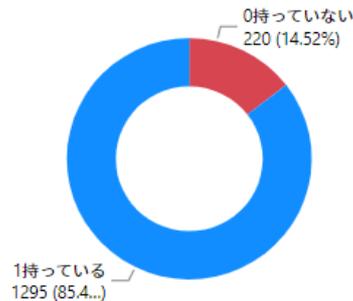


Q2. タブレットの所持状況

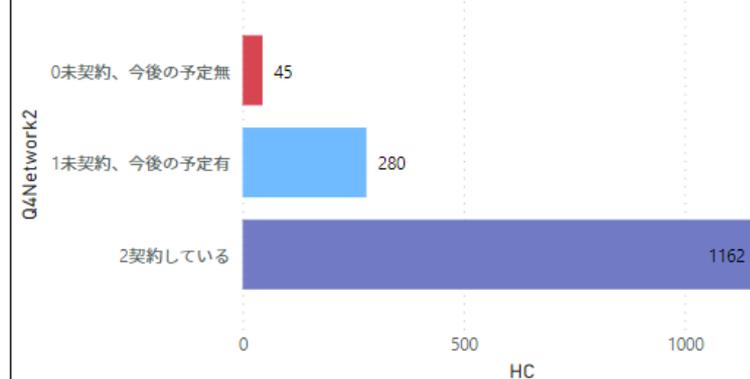


1550
HC

Q3. PCの所持状況



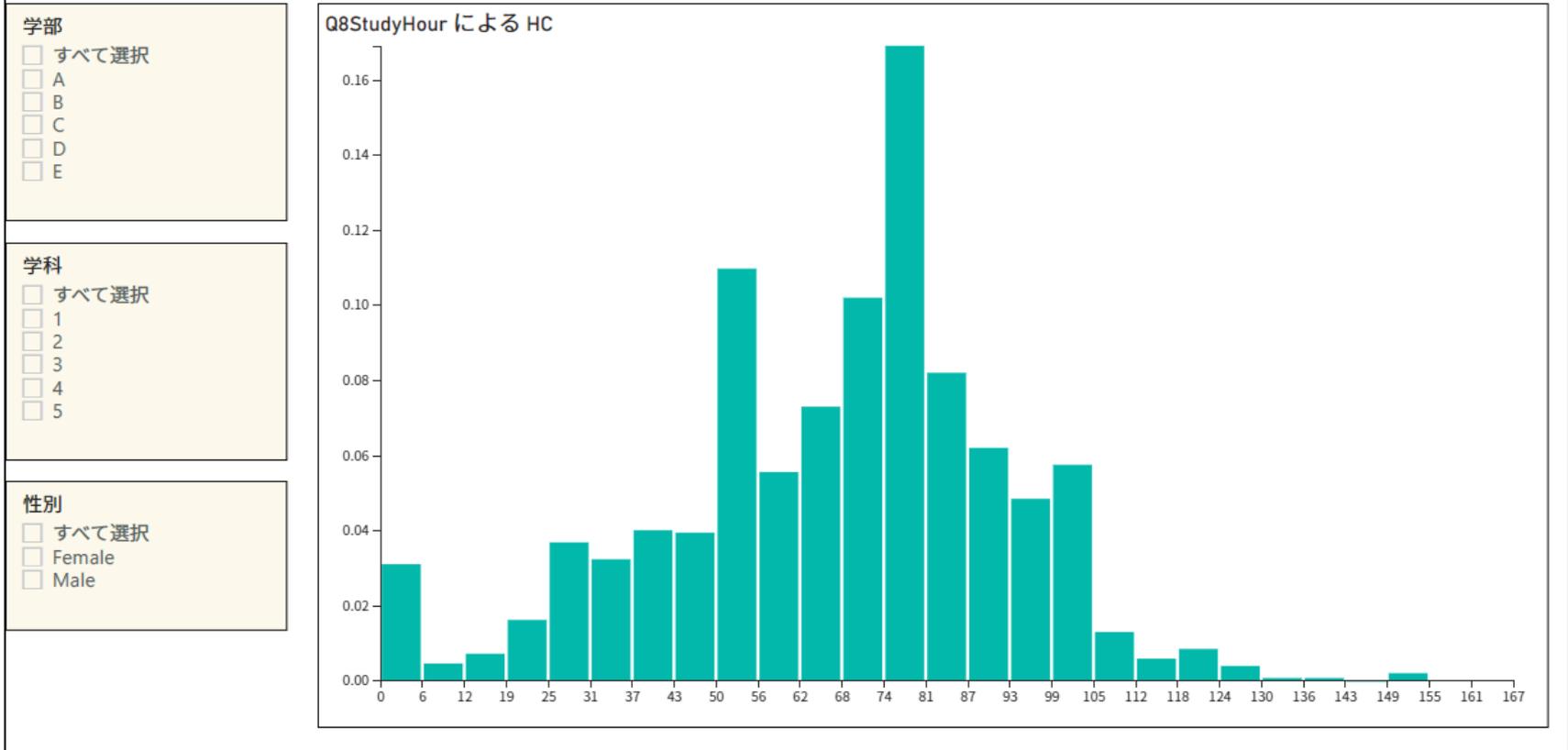
Q4. ネットワークの契約状況



Power BIレポートの例 (2)

IR大学「オンライン授業に関するアンケート調査」

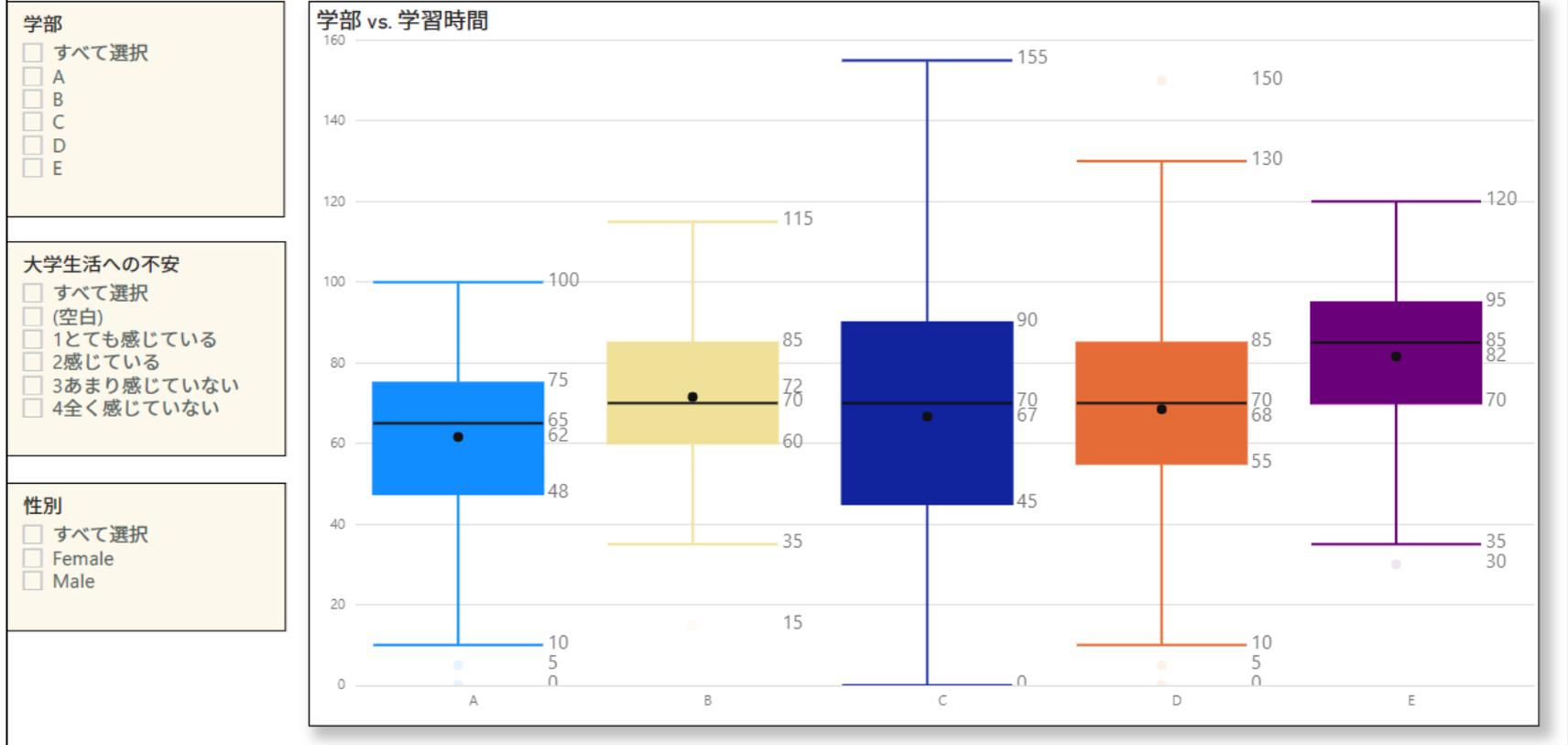
対象：2020年度入学生 (1,550名)



Power BIレポートの例 (3)

IR大学「オンライン授業に関するアンケート調査」

対象：2020年度入学生 (1,550名)



まとめ



まとめ

■ 本プログラムにおける学習がいよいよ**本格化**します。

シラバス科目名	分類コード	授業内容・タイトル	担当	6月	
IR入門	IR101	IR入門	藤原		
IR応用	IR151	IRの発展過程	浅野		
IR応用	IR171	明治大学におけるIR	山本幸一		
IR応用	IR172	山形大学におけるIR	浅野		
IR応用	IR201	IE入門	浅野		
IR応用	IR211	IEとデータ活用	山本幸一		
他大学におけるIR	IR175	TBA (宮崎大学)	田中		
他大学におけるIR	IR175	TBA (帯広畜産大学)	岡部		
他大学におけるIR	IR175	TBA (鶴見大学)	北原		6月18日
他大学におけるIR	IR175	TBA (淑徳大学)	荒木		
他大学におけるIR	IR175	TBA (九州工業大学)	山本鉦		
DMA入門	DMA101	DMA環境の構築	藤原		
DMA入門	DMA102	Rによるデータマネジメント入門	藤原		
DMA入門	DMA102	RによるDMの実践	藤原		
DMA入門	DMA102	Power QueryによるDM例	鈴木		
DMA入門	DMA103	Microsoft Forms入門	山本幸一		
DMA入門	DMA103	Diverging Stacked Bar Chart (Excel)	山本幸一		
DMA入門	DMA103	Power BI入門	藤原		
DMA入門	DMA103	FormsとPower BIの連携	鈴木		
DMA入門	DMA103	アンケート作成の基本	藤原		
DMA入門	DMA103	Rによるデータの可視化	藤原・田中		
DMA入門	DMA103	共分散比を用いた入試分析	岡部		
DMA入門	DMA104	プレゼンテーション発表会	藤原、他		発表会2
DMA入門	DMA104	プレゼン発表会：分析例	田中		
DMA入門	DMA104	Executive Summary	藤原		
DMA入門	DMA104	R Markdownによるレポート作成	田中		

- この**臨時**講義は、DMA系の学習における「**最終完成図**」を共有することを目的に作成しました

(岡部先生の提案により実現)

- みなさんで助け合いながら一緒に勉強していきましょう！

- 我々も勉強しています
(Learning by Teaching)

THANK YOU!

ANY QUESTIONS, COMMENTS OR SUGGESTIONS?

藤原 宏司 | Koji Fujiwara, Ph.D.

kfujiwara@cc.yamagata-u.ac.jp

