Co-funded by
the European Union

Storytelling with Data.

Project Reference No. KA220-HED-57DAE8A7 - 000154019
Grant Agreement: 2023-1-IT02-KA220-HED-
Document Title: Storytelling with Data
Relevant Work package: 3
Nature: Report
Dissemination Level: Restricted to a group specified by the consortium (including the Commission Services)
Document version: 1.0
Date: 31/08/2025
Authors: Georgios Tsaples (IDS), Christos Kassimeris (EUC), Francisco Javier Orquín Castrillón (UCAM), Carmen Daniela Quero Calero (UCAM), Stefano Armenia (LCU) Andra Burlacu (Sydic) Camillo Carlini (Briswa Italia), Jason Papathanasiou (UOM), Panagiota Digkoglou (UOM), Georgios Aretoulis (UOM), , Dimitra Manou (UOM)
Commented/Feedback: All partners
Document description: The objective of this report is to serve both as a textbook and a practical guide on how to use data effectively in addressing social problems and decision-making contexts. To meet this objective, the report is structured around two main pillars. Section 2 focuses on the correct use of standard graphs, outlining how different types of data can be represented clearly and appropriately. Section 3 builds on this foundation by demonstrating how these graphs can be transformed into storytelling devices that enhance understanding and engagement.

Co-funded by
the European Union

The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Table of Contents

List of Figures

Executive Summary

The objective of this report is to serve both as a textbook and a practical guide on how to use data effectively in addressing social problems and decision-making contexts.

To meet this objective, the report is structured around two main pillars. Section 2 focuses on the correct use of standard graphs, outlining how different types of data can be represented clearly and appropriately. Section 3 builds on this foundation by demonstrating how these graphs can be transformed into storytelling devices that enhance understanding and engagement.

The data is not only used to demonstrate how to present technical knowledge to policy makers, but reveals interesting insights as well. The analysis shows that respondents—mostly young students—express optimism toward institutions and programs. They trust their universities, view sport as a tool of integration, and expect inclusion to be a shared norm.

Yet this optimism does not erase lived realities. One in five students reported direct discrimination, almost none received assistance, and incidents concentrate in precisely the spaces—education, public life, and sport—that are most trusted. Especially sports, which hold a high symbolic value as anti-racist institutions, remain an active site of discrimination. Thus, more robust mechanisms and frameworks that take into account the demographic context are necessary to increase trust and transform it into effective anti-racism policies.

Introduction

The rise of big data has created many challenges along with the vast opportunities that it has provided. The scale and velocity of contemporary datasets bring problems of understanding, poor data quality, incorrect integration, costs, verification, and the additional burden of handling information in real time (Lotfi et al., 2023). Managing, curating, processing, and analyzing such amounts of information is therefore not only a technical challenge but also a managerial one. Analytics technologies can process complexity, but managers still face significant difficulties in interpreting the output in ways that are useful to their decisions or decision-making processes (Bumblauskas et al., 2017). Research on how the quality of data interpretation shapes decision-making remains scarce, leaving a gap between data availability and decision effectiveness (Boldosova and Luoto, 2020).

In such a context, storytelling has emerged as a distinct response to the limitations of data-heavy approaches. Traditional reports that rely solely on charts, tables, and numbers often fail to connect with readers, particularly those without strong technical expertise (Knaflic, 2015). Data storytelling goes further: it uses language and communicative abilities to enhance visualizations, enabling audiences to understand where data comes from, why it matters, and what insights it offers (Echeverria et al., 2018). This dual role of translating complexity while maintaining accuracy, has made storytelling a crucial tool for data-driven environments.

Visual data stories combine narrative text with charts, maps, or diagrams, creating reports that are coherent, engaging, and contextually rich (Hullman and Diakopoulos, 2011; Otten et al., 2015). These stories help uncover correlations or causal relationships, highlight trends, and reveal hidden insights that might otherwise be lost in technical outputs (Islam et al., 2024). Because well-crafted stories integrate both numbers and context, they can make data accessible even to audiences with limited technical expertise or data literacy. Kelliher and Slaney (2012) suggest that data stories can serve different functions: to inform by making or refuting claims, to explain by revealing deeper insights, to persuade by stressing the importance of findings, or even to entertain, sustaining engagement and curiosity (Lund, 2022).

Storytelling also operates as a sensemaking tool. Faced with overwhelming volumes of numbers, charts, and bars, readers often struggle to extract meaning. Storytelling transforms raw figures into easily understood insights, connecting analytical outputs and human understanding (Gershon and Page, 2001; Yang, 2013). Effective infographics draw on principles of psychology, usability, graphic design, and statistics to reduce barriers such as limited time or information overload, thus, the combination of multiple forms of visualization into a larger narrative composition, allows audiences to connect individual measures with broader systemic patterns (Otten et al., 2015).

The effectiveness of this approach is not only cognitive but also emotional. Stories that employ surprise can trigger questions, interest, and personal engagement, which in turn sustain learning and discovery. In this sense, stories are not mere tools for communication but part of the broader human process of discovery and growth (Matei and Hunter, 2021).

The challenge is sharper when dealing with social data. Unlike clean numerical datasets, social data is often ambiguous and fragmented or even biased and with missing values. Storytelling provides a means to reduce this ambiguity, to situate data within its context, and to offer a specific interpretation that can guide action (Behesti et al., 2020). However, this requires careful preparation by organizing the raw data, contextualizing it, discovering connected events and entities, and finally presenting it in interactive formats that allow exploration by end-users. Without this preparation, the cost and effort of consuming complex datasets can render them inaccessible, particularly for resource-constrained organizations such as non-profits that rely on open data for goals like community development or policy advocacy (Voida et al., 2011; Erete et al., 2016).

In sum, storytelling with data is more than decoration around charts. It is a communicative practice that integrates technical visualization with narrative sensemaking and it responds directly to the drawbacks of big data with the purpose of enhancing human comprehension. Whether through infographics, interactive visualizations, or narrative reports, data storytelling creates the conditions for knowledge transfer, persuasion, and action in contexts where raw data alone cannot suffice (Shi et al., 2020; Lotfi et al., 2023).

The objective of this report is to serve both as a textbook and a practical guide on how to use data effectively in addressing social problems and decision-making contexts. To meet this objective, the report is structured around two main pillars. Section 2 focuses on the correct use of standard graphs, outlining how different types of data can be represented clearly and appropriately. Section 3 builds on this foundation by demonstrating how these graphs can be transformed into storytelling devices that enhance understanding and engagement. This dual approach is reinforced by the extra material that was created in the context of the current deliverable. In this link, the users can find all the code of Section 2 on the correct use of graphs. They can download the file and run it (even with their own data) locally or they can open it with Google Colab and run the code online. In this link, the users can find the Conclusions sections of the current deliverable transformed in an .html file; they can download it and open it in a browser, where they will see a document which serves as a policy brief and a teaching guide on infographics.

Correct Graphs for different types of data

For this section, we will use the questionnaire that was designed in the context of the current project and especially the responses that were collected from various countries around Europe. The objective of this section is to provide guidelines, along with practical tools/code on how to graph different types of data.

The questionnaire contains different types of data: nominal categories (e.g., gender), binary/ternary answers (yes/no/not sure), ordinal Likert scales (1–5 effectiveness), numeric variables (age), and paired relationships (age vs effectiveness). Each type must be plotted on a specific graph that respects its structure.

Nominal Categories

Nominal categories are labels that classify people or things into groups without any particular order. Common examples are gender, country of residence, or type of occupation. The main characteristic of these categories is that they are "different," but not ranked.

It would be best to use a bar chart for nominal data, because a bar chart makes it easy to see how many people fall into each category and to compare categories side by side, while avoiding unnecessary complexity while still showing the overall distribution.

The code below is written in the programming language of Python (van Rossum & Drake, 2009) and it can be copied and paste to any python compiler, and by changing the data the user can generate their own graphs of nominal data.

In the current example, the graph illustrates the gender distribution of the questionnaire respondents.

```
import matplotlib.pyplot as plt
# Example data: gender distribution
gender_counts = {"Male": 115, "Female": 19, "Non-binary": 2}
labels = list(gender_counts.keys())
values = list(gender_counts.values())
plt.figure()
plt.bar(labels, values)
plt.title("Gender distribution")
plt.ylabel("Number of respondents")
plt.xlabel("Gender")
plt.show()
```

The graph on Figure 1 illustrates that 115 of the responses came from men, 19 from women and only 2 who characterized themselves as non binary.

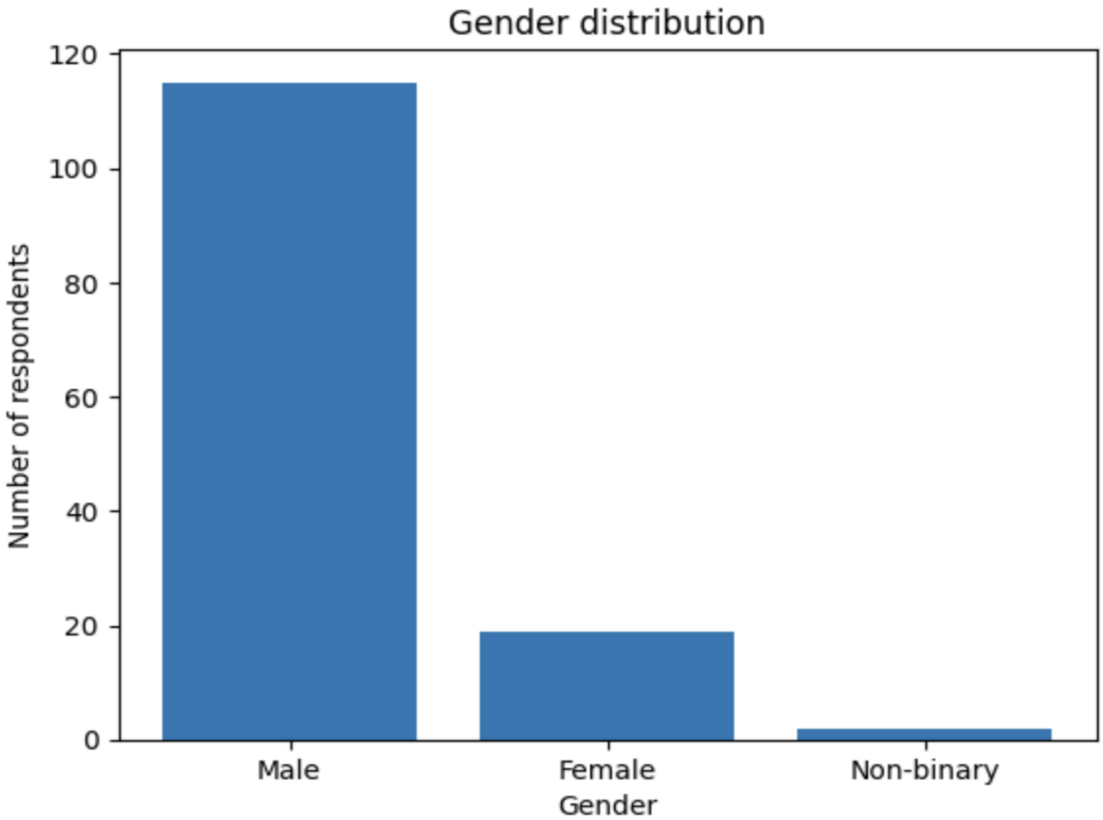


Figure 1 Gender Distribution

Binary data

Sometimes a question offers only two or three options, such as “Yes/No” or “Yes/No/Not sure.” These are categorical as well, but here the order often matters a little more, since “Yes” is usually the most direct answer, followed by “No,” and then perhaps “Not sure.”.

In such a context, it would be again best to use a bar chart, but here the important point is to place the categories in a logical order (for example, Yes → Not sure → No), which makes the answers easier to follow and avoids confusion.

```
Import matplotlib.pyplot as plt
# Example data: sports as an effective anti-racism tool
sport_effective = {"Yes": 100, "Not sure": 20, "No": 13}
labels = ["Yes", "Not sure", "No"]
values = [sport_effective.get(k, 0) for k in labels]
plt.figure()
plt.bar(labels, values)
plt.title("Sport as an effective anti-racism tool")
plt.ylabel("Number of respondents")
plt.xlabel("Response")
plt.show()
```

The graph on Figure 2 shows that 100 respondents consider sports as an effective tool against racism, with 20 not being sure of its effectiveness and only 13 declaring that they do not consider it effective as an anti-racism tool.

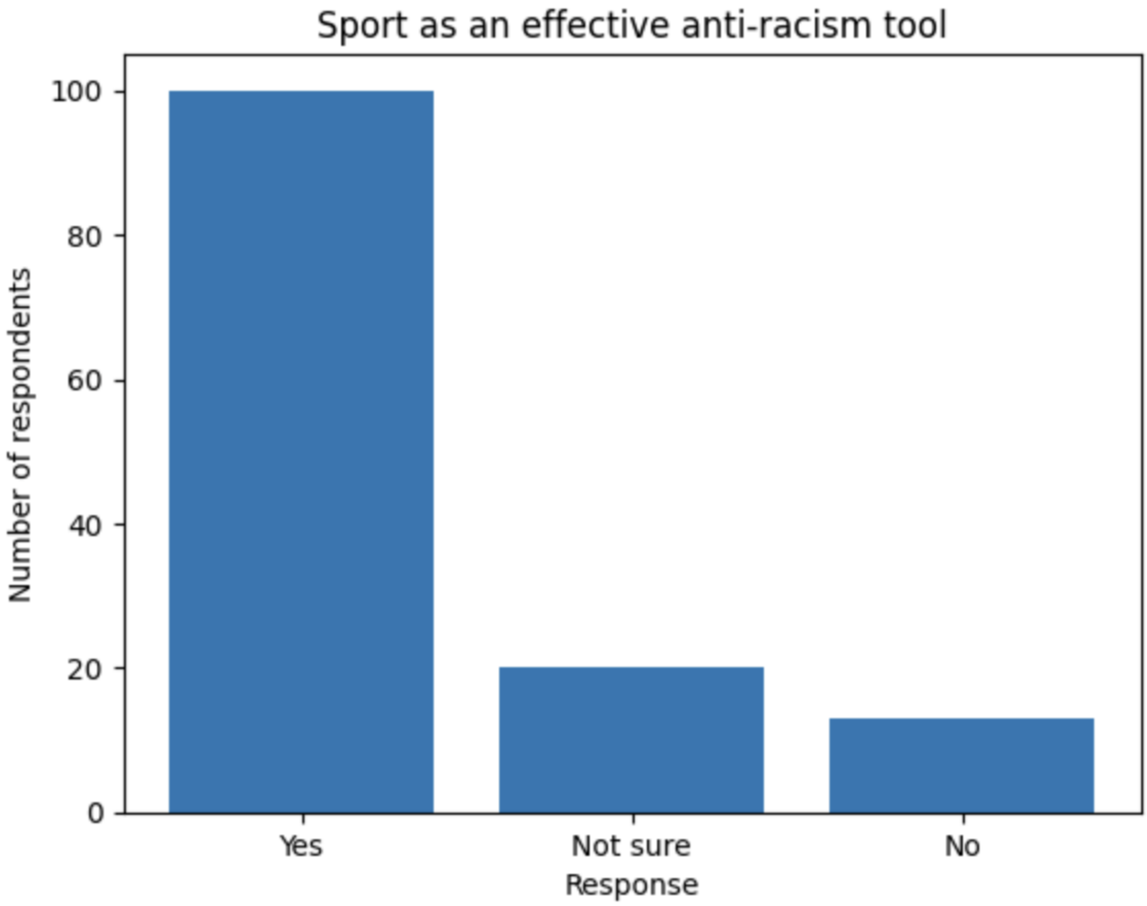


Figure 2 Responses to the question of the effectiveness of sports as an anti-racism tool

Ordinal data (Likert scale)

Ordinal data refers to categories that do have a meaningful order. Likert scales are the most common example: respondents may rate a program from 1 (very ineffective) to 5 (very effective). Each number means "more" or "less," but the gap between them is not always equal; going from 1 to 2 is not necessarily the same as moving from 4 to 5.

In such a case, it would be best to use an ordered bar chart that shows how many people chose each point on the scale, because this allows readers to see whether opinions cluster around the middle, or whether they lean toward the positive or negative side. Showing the full distribution is better than reducing everything to an average.

```
Import matplotlib.pyplot as plt
#Example data: effectiveness ratings (1-5)
likert_counts = {1: 14, 2: 7, 3: 60, 4: 29, 5: 27}
levels = [1, 2, 3, 4, 5]
values = [likert_counts.get(l, 0) for l in levels]
plt.figure()
plt.bar([str(l) for l in levels], values)
plt.title("Effectiveness of educational programs (1-5)")
plt.ylabel("Number of respondents")
plt.xlabel("Likert level")
plt.show()
```

Figure 3 below shows how effective the respondents consider educational programs as a means to fight racism.

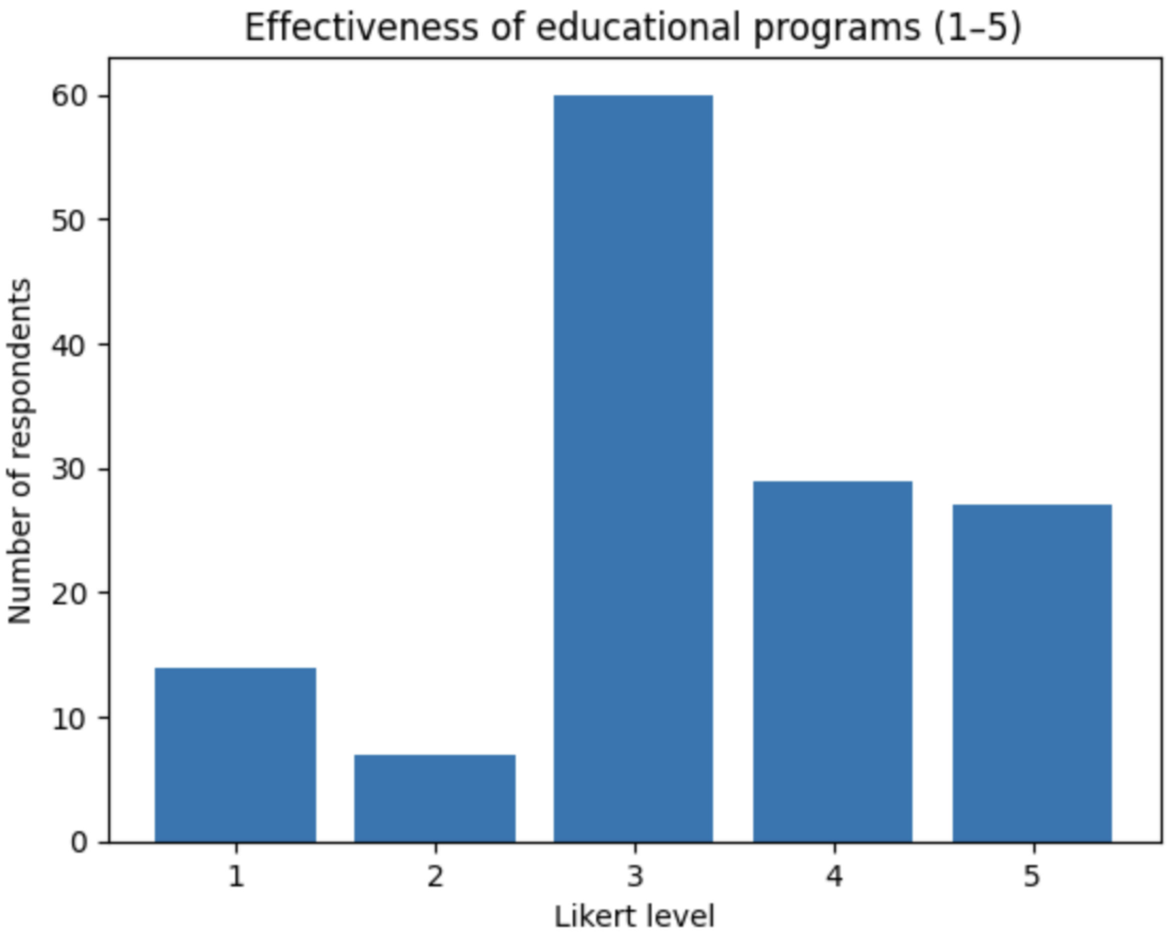


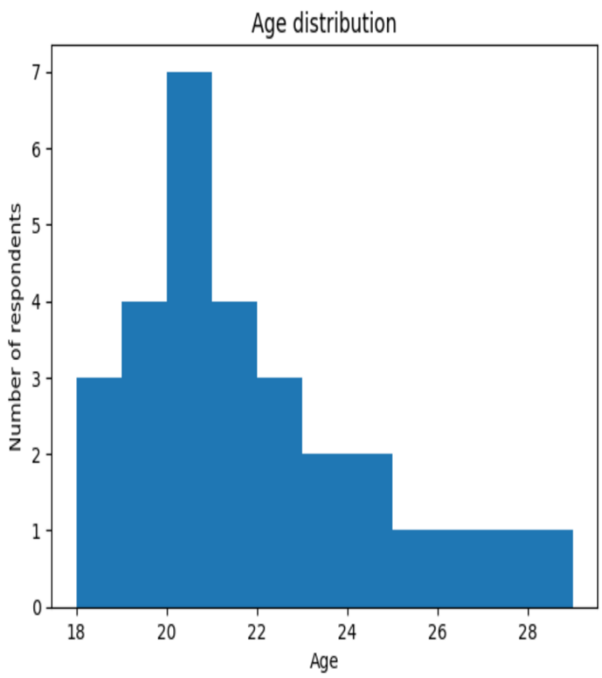
Figure 3 Effectiveness of educational programs

Numerical Data

Numerical data are values that can be measured and ordered naturally, such as age or income. Unlike categories, numbers allow us to talk about averages, ranges, and distributions. For numeric data it would be best to use a histogram if you want to see the shape of the distribution (are most respondents young? is there a wide spread?), and a box plot if someone wants to summarize the spread (minimum, maximum, quartiles, and median). Both highlight different aspects of the same data.

```
Import matplotlib.pyplot as plt
# Example data: ages sample from the survey
Ages = [18,19,20,21,22,23,24,25,26,27,28,
  18,18,19,20,20,21,22,22,23,24,20,19,21,20,20,21,19,20]
# Histogram
plt.figure()
plt.hist(ages, bins=range(18, 30))
plt.title("Age distribution")
plt.xlabel("Age")
plt.ylabel("Number of respondents")
plt.show()
# Box plot
plt.figure()
plt.boxplot(ages, vert=True)
plt.title("Age (box plot)")
plt.ylabel("Age")
plt.show()
```

Figure 4 below illustrates both a histography and a boxplot of a sample of ages that were reported in the responses on the questionnaire.



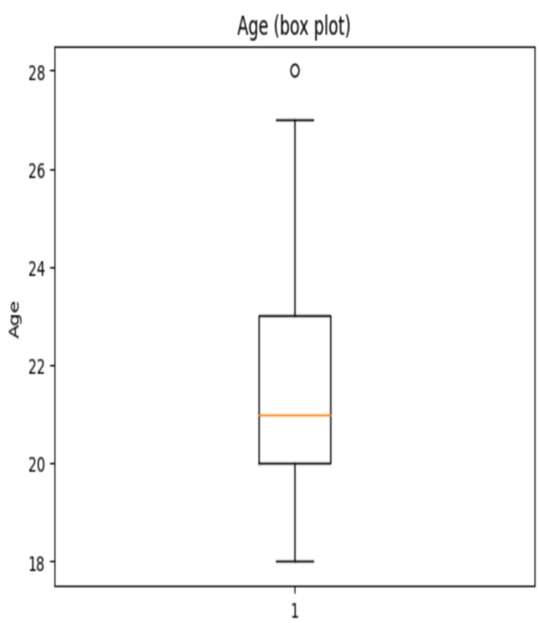


Figure 4 Histogram (Left) and boxplot (right) of the age of the respondents

Relationships between two variables

There are cases where we would like to investigate how two variables relate to each other. For example, whether age has any influence on how students rated the effectiveness of a program. In the particular case, one variable is numeric (age), the other is ordinal (effectiveness).
In such a context, it would be best to use a scatter plot: Each dot represents one person's answers on both questions. This makes it possible to see whether older or younger respondents tend to rate differently, without forcing an artificial summary.

```
import matplotlib.pyplot as plt
# Example paired values: (Age, Effectiveness)
age_eff_pairs = [
 [19,5],[23,3],[22,2],[22,1],[18,4],[21,3],[19,3],[25,2],[20,3],[20,5]
]
ages = [p[0] for p in age_eff_pairs]
```

Hence, in figure 5 we can see how age correlates with the answer on the effectiveness of an educational program as an anti-racist tool.

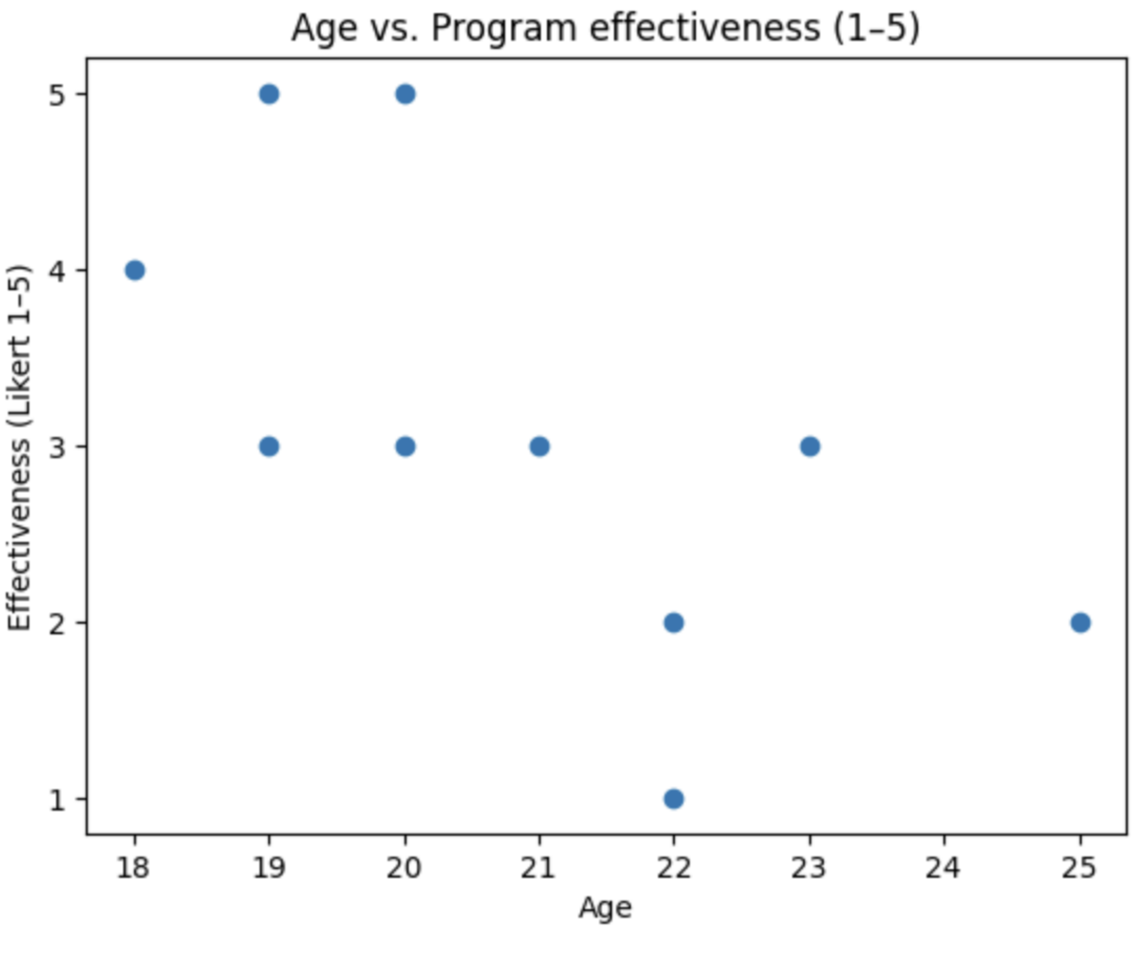


Figure 5 Age and effectiveness of educational programs

Multi-response questions

Some questions allow multiple answers, for example: “In what context have you experienced racial discrimination? (select all that apply).” Here, one respondent may choose “University” and “Online,” another might choose “Workplace” and “Public Transport.”
It would be best to count how often each context appears across all responses and then use a horizontal bar chart. Horizontal bars are easier to read when labels are long, as is often the case with these questions.

```
import matplotlib.pyplot as plt
from collections import Counter
multi_responses = [
"Public Transport, Workplace",
"University, Public Transport",
"Workplace",
"Sports Club, Public Spaces",
"Public Spaces, University",
"Public Spaces, University"

]
labels_counter = Counter()
for resp in multi_responses:
    parts = [p.strip() for p in resp.split(",") if p.strip()]
    labels_counter.update(parts)
labels = list(labels_counter.keys())
values = [labels_counter[k] for k in labels]
plt.figure()
plt.barh(labels, values)
plt.title("Contexts reported (select-all question)")
plt.xlabel("Mentions")
plt.ylabel("Context")
plt.show()
```

Figure 6 illustrates how an horizontal bar chart looks like for the question “In what context have you experienced racial discrimination?” and a sample of the responses.

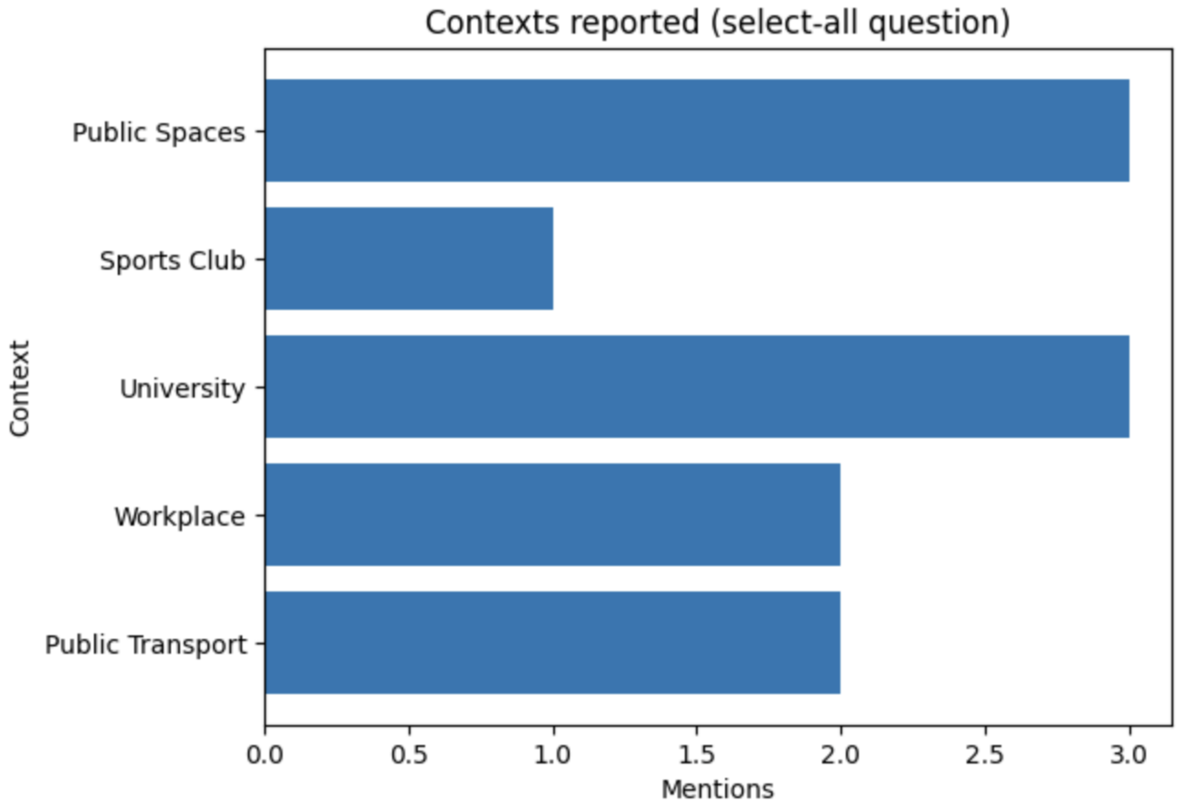


Figure 6 Horizontal bar chart for multi-response questions

Time-based data

If a dataset contains dates or timestamps (for instance, when people submitted the questionnaire), the focus is usually on how the numbers change over time. In such a case, it would be best to use a line chart, since it highlights trends and patterns across time intervals. Before plotting, it is helpful to group the data into sensible units (for example, per week).

```
import matplotlib.pyplot as plt
weeks = ["Wk1","Wk2","Wk3","Wk4","Wk5"]
submissions = [12, 25, 19, 31, 22]
plt.figure()
plt.plot(weeks, submissions, marker="o")
plt.title("Weekly form submissions")
plt.xlabel("Week")
plt.ylabel("Count")
plt.show()
```

For example Figure 7 illustrates the number of responses that were submitted for the time period (weeks) that the questionnaire was active.

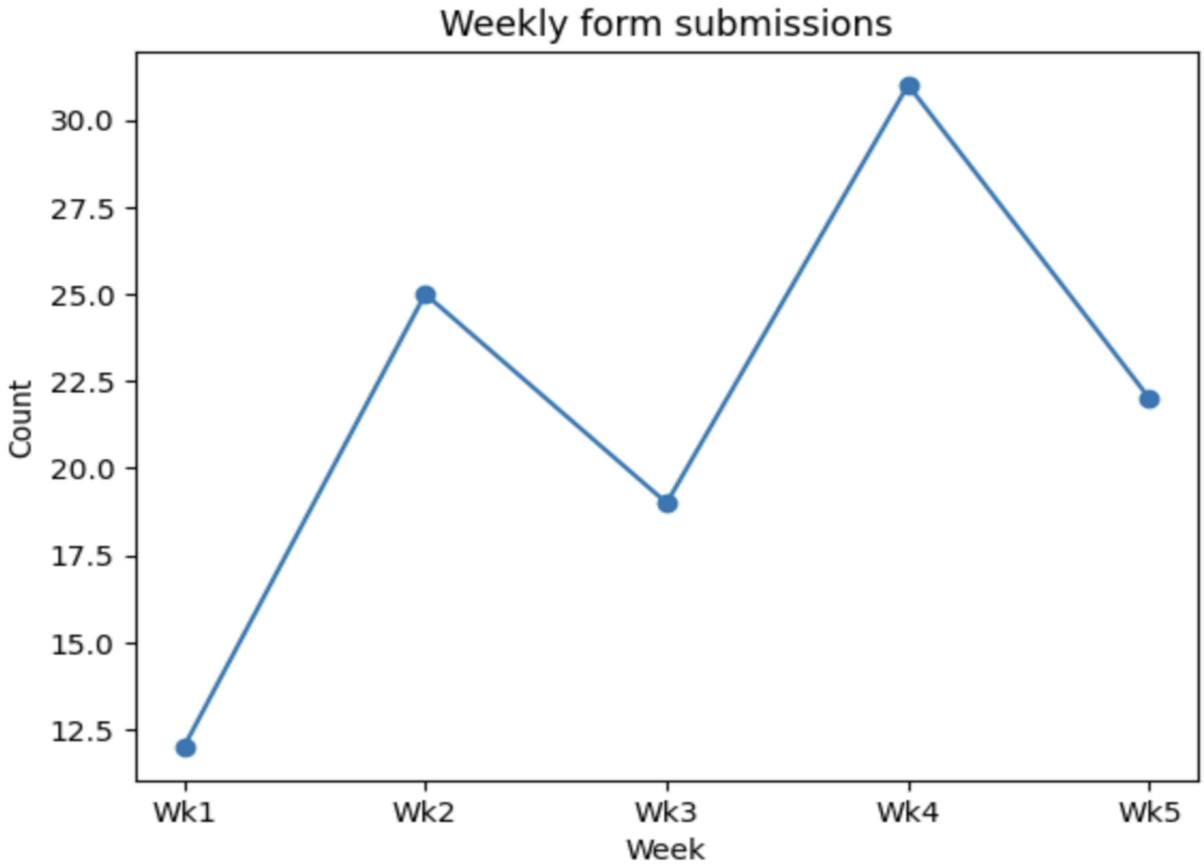


Figure 7 Number of responses per week as a line graph

Consequently, each type of data has its own "best fit" when it comes to visualization. Nominal and binary categories are most clearly shown as bar charts; ordinal scales are best fitted as ordered bars; numeric values in histograms and box plots; relationships in scatter plots; multi-response items call for horizontal bars; and time is best seen through line charts. Choosing the right match between data and graph ensures that the story the data can tell is both accurate and easy to understand. The code that has been used in this section can be found here, and it can be either downloaded or even run in the Google Colab environment.

Storytelling with data

In the previous section, the focus was on how to select the right type of graph for the right type of data. This step is essential because it ensures accuracy and clarity. However, as it was mentioned in the introduction, it is necessary to combine data in order to tell a full story especially for difficult issues such as racism. A story emerges when different pieces of information are brought together, when patterns are connected, and when meaning is drawn from the numbers and words on the page.

This section takes the next step by showing how data can be transformed into a narrative. The aim is not only to present results but to guide the reader through them in a way that highlights contrasts, reveals connections, and explains why the data matters. Storytelling allows us to move from isolated graphs to a coherent picture of racism and discrimination, where the whole becomes greater than the sum of its parts.

We will begin with the survey data collected in the BRISWA 2.0 project, since these responses provide an entry point into the issue. From there, we will move to external numerical indicators, which situate those experiences within broader structural patterns. We will then bring in textual insights from reports and analyses that were performed in the context of the BRISWA 2.0 project, which reveal how racism is discussed, categorized, and sometimes minimized by institutions. Finally, we will look at the programs and pedagogical approaches identified by the project, which represent efforts to respond to the challenges uncovered by the data.

The section concludes by bringing these streams together into a single story. Survey data, numerical indicators, textual insights, and program analysis will be combined through interactive and visually engaging formats. The result will not just be a collection of charts but a narrative that shows how racism manifests, how it is experienced, how it is framed, and how it can be addressed. In this way, the data does not remain static; it becomes a story that informs, persuades, and inspires action.

Questionnaire data

Similar to the previous section, we will begin with the BRISWA 2.0 questionnaire.

The first step in understanding the survey results is to look at the demographic profile of respondents. Age and gender are especially relevant, as they shape experiences of racism and influence perceptions of inclusion. The dataset shows a young population, largely concentrated between 18 and 24 years old, and a strong gender imbalance, with male respondents far outnumbering female respondents.

```
import matplotlib.pyplot as plt
# --- Inline data from the BRISWA survey (exact number)---
# Gender distribution (exact values from the survey)
gender_counts = {
    "Male": 115,
    "Female": 19,
    "Non-binary": 2,
    "Male, Female": 1
}
# Full list of recorded ages (n = 135)
Ages = [
    24,20,22,22,18,21,19,25,20,20,21,19,20,18,24,22,19,23,22,21,
    21,21,20,20,19,18,19,21,20,22,21,20,20,20,20,20,20,19,20,20,
    20,20,20,23,26,25,20,21,22,20,23,19,20,22,20,19,20,20,19,20,
    19,20,21,20,20,21,20,18,21,19,20,19,20,20,23,22,21,22,20,20,
    20,19,20,20,20,19,20,22,20,20,20,20,19,20,20,21,20,22,20,20,
    19,20,22,21,21,23,21,20,20,20,20,22,19,19,21,22,24,19,19,22,
    28,20,20,21,20,19,20,22,18,20,22,19,20
]
# --- Create subplots side by side ---
fig, axes = plt.subplots(1, 2, figsize=(12, 5))
# Plot 1: Gender distribution
axes[0].bar(gender_counts.keys(), gender_counts.values(), color="skyblue")
axes[0].set_title("Gender distribution")
axes[0].set_ylabel("Number of respondents")
axes[0].set_xlabel("Gender")
# Plot 2: Age distribution
axes[1].hist(ages, bins=range(18, 30), color="salmon", edgecolor="black")
axes[1].set_title("Age distribution")
axes[1].set_xlabel("Age")
axes[1].set_ylabel("Number of respondents")
plt.tight_layout()
plt.show()
```

The age histogram confirms a student-heavy sample with a narrow band centered around early adulthood. The gender bar chart shows a strong male skew, plus one multi-marked entry (“Male, Female”) that we kept visible to reflect the raw responses. This uneven demographic profile is important context for the rest of Section 3: attitudes and experiences that will be revealed later may reflect the viewpoints of a predominantly young, male group (Figure 8).

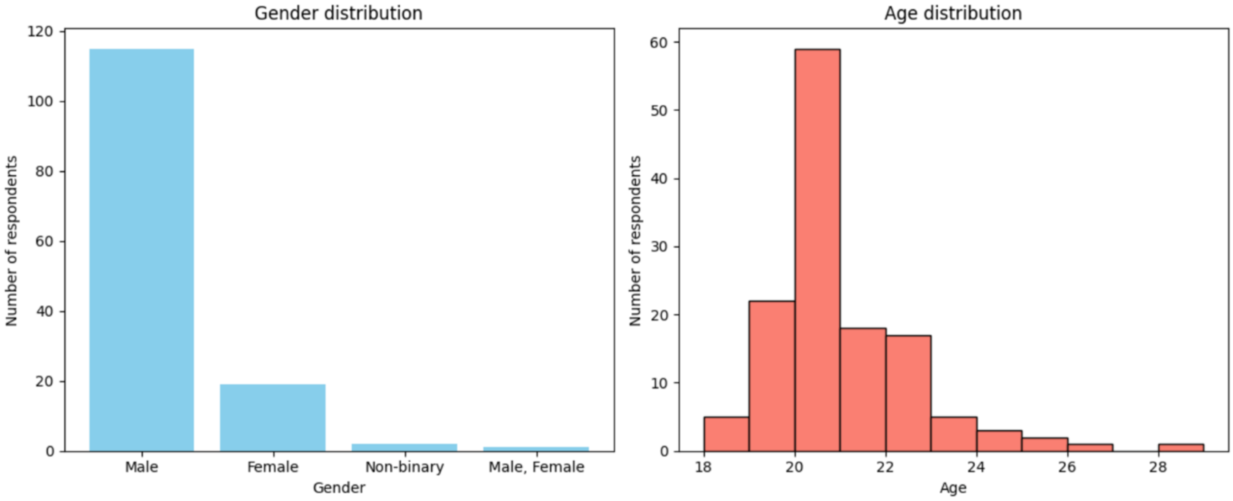


Figure 8 Age and gender distribution for the BRISWA 2.0 dataset

Experiences of discrimination

The next step is to move to “what the respondents have experienced.” The responses show that racial discrimination is not abstract: almost one in five students reported having faced it since arriving in their host country. Among the 137 respondents, 109 answered “No,” 26 answered “Yes,” and two marked both options. Together, this means that 28 students — roughly 20% — described having personally encountered discrimination.

The question of what happens after such an incident reveals an even deeper concern. Of those who reported discrimination, only one respondent (3.6%) received assistance. Hence, for nearly everyone who faced discrimination, there was no support to meet their needs (or support was never asked for).

```
import matplotlib.pyplot as plt
# --- Inline survey data (exact tallies) ---
discr_experience ={"No": 109, "Yes": 26, "Yes, No": 2}
n_total = sum(discr_experience.values())
n_yes = 26 + 2   # count "Yes" + mixed
# Context percentages among the 28 "Yes"
context_info = [
    "Of those who responded Yes:",
    "57.1% experienced discrimination in a public space",
    "46.5% experienced discrimination in an educational setting",
    "39.3% experienced discrimination in Social media",
    "7.1% experienced discrimination in a football setting"
]
# --- Donut chart ---
fig, ax = plt.subplots(figsize=(8, 6))
sizes = list(discr_experience.values())
labels = list(discr_experience.keys())
colors= ["#6baed6", "#fd8d3c", "#969696"] # blue = No, orange = Yes, gray =
Yes/No
wedges, _ = ax.pie(sizes, startangle=90, colors=colors)
# Donut hole
centre_circle = plt.Circle((0, 0), 0.60, fc="white")
ax.add_artist(centre_circle)
# Title
ax.set_title("Experienced racial discrimination?")
# Central text (slightly smaller)
pct_yes = round(100.0 * n_yes / n_total, 1)
ax.text(0,0.07,f"{pct_yes}%",ha="center",va="center",fontsize=12,weight="bold
")
ax.text(0,-0.06,"reported‘Yes’",ha="center",va="center",fontsize=11)
ax.text(0,-0.20,"3.6% of ‘Yes’ received assistance",
     ha="center", va="center", fontsize=9, color="darkred")
# Legend box positioned near the "Yes" slice (orange, usually right side)
legend_text = "\n".join(context_info)
ax.text(1.2, 0, legend_text,
     ha="left", va="center",
     fontsize=10,
     bbox=dict(boxstyle="round,pad=0.5", fc="white", ec="0.5"))
ax.set_aspect('equal')
plt.tight_layout()
plt.show()
```

The donut chart below (Figure 9) visualizes these answers in a compact way. The orange segment corresponds to those who reported discrimination with the text in the center gives the percentage of “Yes” responses and the fraction that received assistance. To the side, we break down the contexts in which discrimination occurred. These are not evenly distributed: more than half of those affected mentioned public spaces, almost half mentioned educational settings, and more than a third pointed to social media, while football and other contexts were mentioned less frequently.
The combination of these responses provides a compact insight into the perspectives of the respondents: discrimination is a reality for a minority of students, help is either almost absent or not

asked for, and the places where incidents occur are concentrated in everyday public and institutional life.

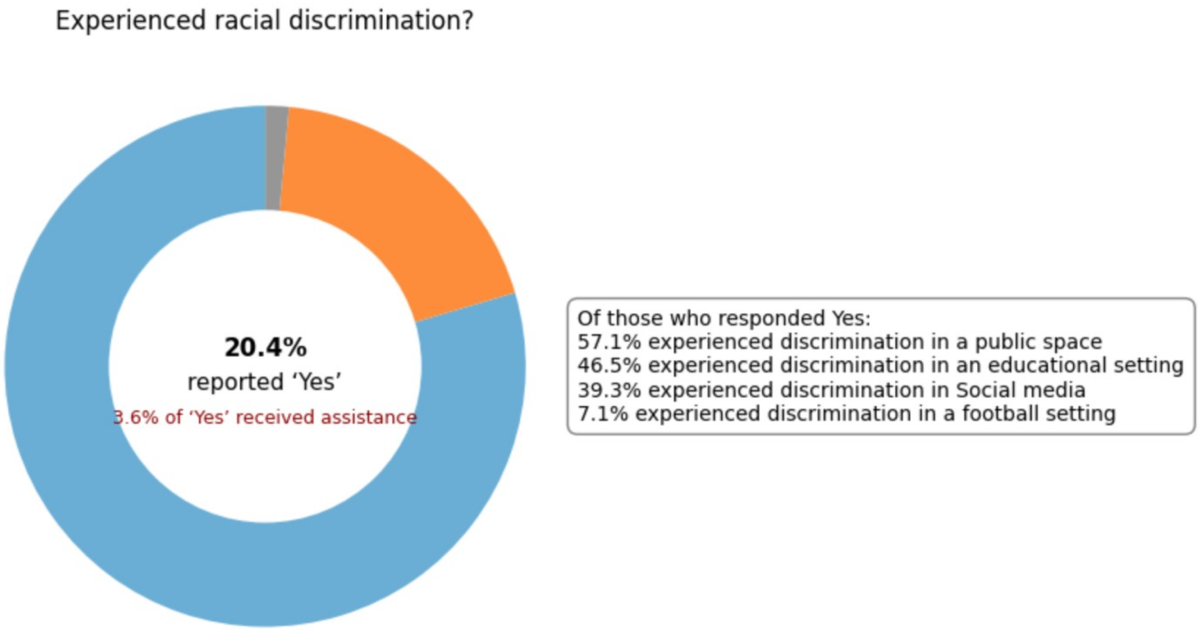


Figure 9 Discrimination reported by 20% of students, with assistance nearly absent (3.6% of cases). Incidents are concentrated in public spaces, education, and social media.

Perceptions of institutions

Experiences of racism do not occur in a vacuum. They are shaped by the response of institutions, which act as the agents responsible for fostering inclusion. If students perceive that their institutions are unsupportive or ineffective, incidents of discrimination risk being overlooked or minimized. The BRISWA 2.0 contained three, relevant questions:

1. How students rated the support received from their educational institutions for integration.
2. Whether they thought the programmes at their university adequately addressed the issue of racism.
3. How effective they considered their university to be in promoting inclusion and respect for diversity.

To compare these three dimensions, responses were grouped into negative (ratings of 1–2 or "No"), neutral (rating of 3 or "Not sure"), and positive (ratings of 4–5 or "Yes"). The diverging stacked bar chart in Figure 10 allows all three questions to be read on the same scale, highlighting the overall balance between skepticism, uncertainty, and trust in institutions.

```
import matplotlib.pyplot as plt
from matplotlib.patches import Patch
# --- Inline counts from the questionnaire---
support_counts = {1: 2, 2: 4, 3: 29, 4: 33, 5: 69}
prog_counts_raw={'No':17,'No,Not sure':1,'Not sure':32,'Yes':86,'Yes, No':1}
effective_counts = {1: 1, 2: 4, 3: 26, 4: 51, 5: 55}
def likert_to_buckets(counts):
    total = sum(counts.values())
    neg = counts.get(1, 0) + counts.get(2, 0)
    neu = counts.get(3, 0)
    pos = counts.get(4, 0) + counts.get(5, 0)
    return neg, neu, pos, total
def yns_to_buckets(counts):
    neg = counts.get('No', 0)
    pos = counts.get('Yes', 0)
    neu = counts.get('Not sure', 0) + counts.get('Yes, No', 0) + counts.get('No, Not sure', 0)
    total = neg + neu + pos
    return neg, neu, pos, total
# Bucketize and convert to %
def to_pct(neg, neu, pos, total):
    f = 100.0 / total if total else 0
    return neg*f, neu*f, pos*f
sN, sU, sP = to_pct(*likert_to_buckets(support_counts))
pN, pU, pP = to_pct(*yns_to_buckets(prog_counts_raw))
eN, eU, eP = to_pct(*likert_to_buckets(effective_counts))
labels = [
    "Support for integration",
    "Programmes address racism",
    "Effectiveness on inclusion"
]
Rows = [(sN, sU, sP), (pN, pU, pP), (eN, eU, eP)]
# --- Consistent colors across all questions ---
NEG_COLOR = "#d73027"  # negatve (left)
NEU_COLOR = "#f7f7f7" # neutral (center)
POS_COLOR = "#1a9850" # positive (right)
fig, ax = plt.subplots(figsize=(10, 6))
ypos = list(range(len(rows)))[::-1]
for y, (neg, neu, pos) in zip(ypos, rows):
    # Negative segment (to the left)
    ax.barh(y, -neg, left=0, align='center', color=NEG_COLOR, edgecolor='none')
    # Neutral segment (centered at 0 to the right)
    ax.barh(y, neu, left=0, align='center', color=NEU_COLOR, edgecolor='none')
    # Positive segment (to the right, stacked after neutral)
    ax.barh(y, pos, left=neu, align='center', color=POS_COLOR, edgecolor='none')
# Y labels
ax.set_yticks(ypos)
ax.set_yticklabels(labels)
# Symmetric x-limits and axis labels
ax.set_xlim(-100, 100)
ax.set_xlabel("Share of respondents (%)")
ax.set_xticks([-100, -75, -50, -25, 0, 25, 50, 75, 100])
# Center line
ax.axvline(0, linewidth=1, color="#999999")
# Annotations (skip tiny segments to avoid clutter)
def annotate(x_left, width, y, text):
    If abs(width) < 4: # <4% too small to label cleanly
       return
    ax.text(x_left + width/2, y, text, va='center', ha='center', fontsize=9)
for y, (neg, neu, pos) in zip(ypos, rows):
     annotate(-neg, neg, y, f"{neg:.0f}%")
     annotate(0, neu, y, f"{neu:.0f}%")
     annotate(neu, pos, y, f"{pos:.0f}%")
# Legend with the SAME colors
Handles = [
    Patch(facecolor=NEG_COLOR, edgecolor='none', label="Negative (1–2 or No)"),
    Patch(facecolor=NEU_COLOR, edgecolor='none', label="Neutral (3 or Not sure / mixed)"),
    Patch(facecolor=POS_COLOR, edgecolor='none', label="Positive (4–5 or Yes)")
]
Legend = ax.legend(handles=handles, loc="upper center", ncol=3,bbox_to_anchor=(0.5,1.12),
frameon=False)
plt.tight_layout(rect=(0, 0.03, 1, 1))
plt.show()
```

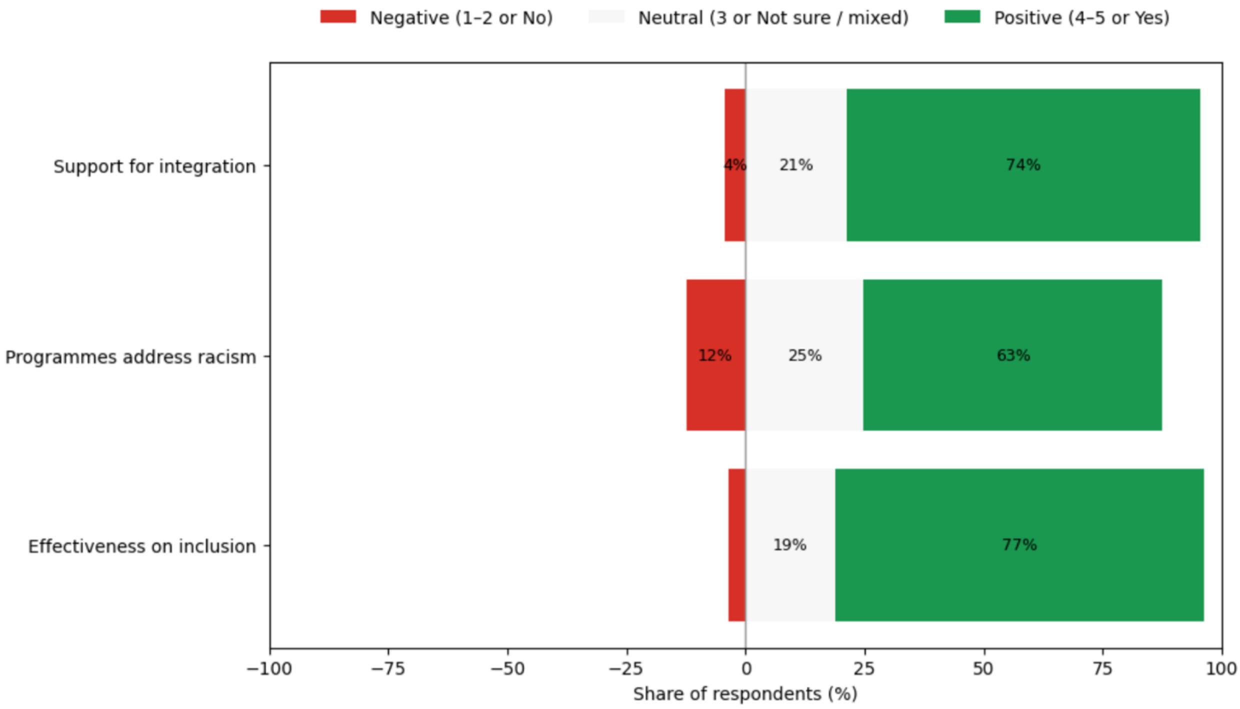


Figure 10 Perceptions of institutions. Most students gave positive ratings for support, programme content, and inclusion, but around one-fifth remained neutral, suggesting mixed or uncertain experiences.

The role of sports

Students overwhelmingly believe that sport can act as a vehicle for integration and respect across cultures: 92% of survey respondents answered "Yes" when asked. This optimism reflects a view of sport as a natural bridge across diversity along with the general young age of the majority of the respondents as was reported in the first graph of the previous section.

At the same time, discourse in wider reports and analyses shows that sport appears alongside themes of *racism, discrimination, harassment, and institutional responsibility*. These keywords highlight that while sport holds promise, it is also a space where problems surface and must be addressed directly.

```
import matplotlib.pyplot as plt
# --- Survey tallies ---
sport_counts = {"Yes": 126, "No": 11}
# -- Thematic keywords from text analysis (top sport-related and adjacent
terms) ---
Keywords = {
    "racism": 45,
    "discrimination": 24,
    "sport": 15,
    "harassment": 13,
    "online": 12,
    "institutional": 18
}
# Left: Survey bar chart
fig, axes = plt.subplots(1, 2, figsize=(12, 5))
# Plot 1: Survey result
axes[0].bar(sport_counts.keys(), sport_counts.values(),
           color=["#1a9850", "#d73027"])
axes[0].set_title("Is sport a tool for integration?")
axes[0].set_ylabel("Number of respondents")
for i, v in enumerate(sport_counts.values()):
    axes[0].text(i, v + 1, str(v), ha="center", va="bottom")
# Plot 2: Keyword frequencies (horizontal bar chart)
kw_items = sorted(keywords.items(), key=lambda x: x[1], reverse=True)
labels = [k.capitalize() for k, v in kw_items]
values = [v for k, v in kw_items]
axes[1].barh(labels, values, color="#6baed6")
axes[1].invert_yaxis()
axes[1].set_title("Themes in wider discourse on racism & sport")
axes[1].set_xlabel("Mentions in reports")
plt.tight_layout()
plt.show()
```

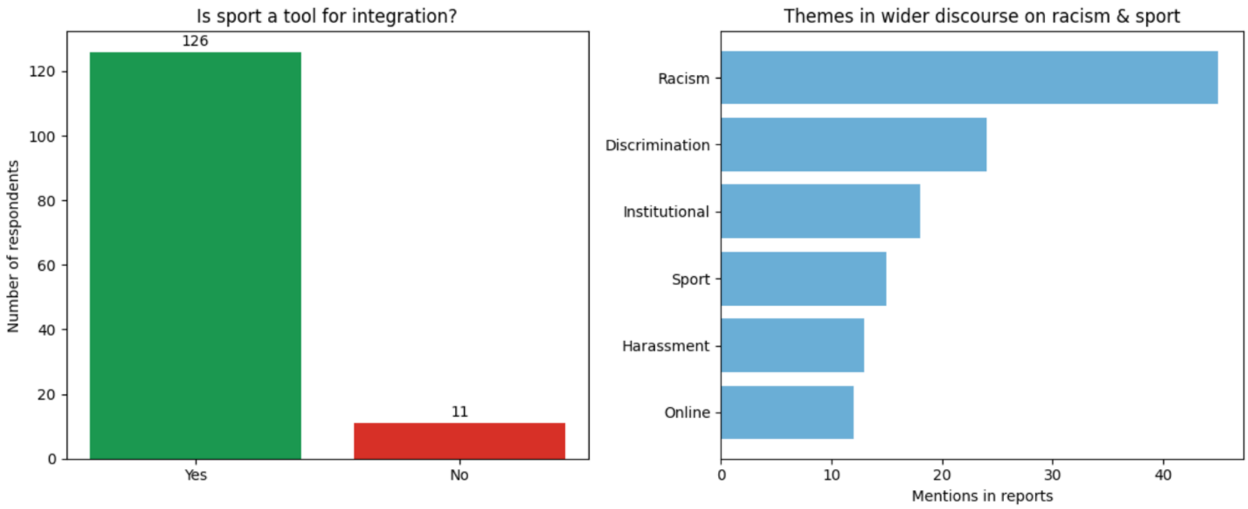


Figure 11 Survey results show that 92% of students view sport as a tool for integration, while thematic analysis of wider reports highlights recurring concerns such as racism, discrimination, harassment, and institutional responsibility.

General Context

The personal experiences in the BRISWA 2.0 Questionnaire are not in a vacuum; rather they can be traced within broader demographic societal realities, including for example migration in Spain. To keep the focus clear, the infographic below highlights: (i) gender composition among newly nationalized people, with the total number of nationality grants noted for context; (ii) an age comparison that illustrates how integration trajectories can differ by group; and (iii) a small set of vulnerability indicators that shape everyday inclusion.

```
import matplotlib.pyplot as plt
# -------------------
# Inline indicators (example: Spain, recent years)
# -------------------
NATIONALITY_GRANTS_2024 = 221_805
WOMEN_SHARE_NEW_NATIONALS = 57  # %
MEN_SHARE_NEW_NATIONALS   = 100 - WOMEN_SHARE_NEW_NATIONALS  # 43%
AVG_AGE_ALL_NEW_NATIONALS = 34
AVG_AGE_MOROCCAN_NATIONALS = 20
PCT_LONG_DURATION_PERMITS = 53.0
PCT_FOREIGN_BORN_UNDER_30 = 35.8
PCT_MIGRANT_WOMEN_SEVERE_POVERTY = 16.7
# -------------------
# Figure layout: reverse pyramid (2 on top, 1 below spanning both)
# -------------------
fig = plt.figure(figsize=(12, 8))
gs = fig.add_gridspec(2, 2, height_ratios=[1, 1.1])
# TOP-LEFT: Donut (% women among new nationals) with central legend incl. nationality
grants
ax0 = fig.add_subplot(gs[0, 0])
sizes = [WOMEN_SHARE_NEW_NATIONALS, MEN_SHARE_NEW_NATIONALS]
labels = ["Women", "Men"]
# Donut (pie with width)
wedges, _ = ax0.pie(sizes, startangle=90)
# hole
ax0.add_artist(plt.Circle((0, 0), 0.60, fc="white"))
ax0.set_title("% Women among new nationals")
# central legend-style text
ax0.text(0, 0.10, f"{WOMEN_SHARE_NEW_NATIONALS}%", ha="center", va="center", fontsize=12,
weight="bold")
ax0.text(0, -0.02, "Women", ha="center", va="center", fontsize=10)
ax0.text(0, -0.18, f"Nationality grants (2024):\n{NATIONALITY_GRANTS_2024:,}",
         ha="center", va="center", fontsize=9)
ax0.set_aspect('equal')
# TOP-RIGHT: Age comparison bar chart
ax1 = fig.add_subplot(gs[0, 1])
groups = ["All new nationals", "Moroccan nationals"]
values = [AVG_AGE_ALL_NEW_NATIONALS, AVG_AGE_MOROCCAN_NATIONALS]
ax1.bar(groups, values)
ax1.set_ylabel("Average age (years)")
ax1.set_title("Average age of newly nationalized")
for i, v in enumerate(values):
    ax1.text(i, v + 0.6, str(v), ha="center", va="bottom")
# BOTTOM (spanning both columns): Vulnerabilities horizontal bars
ax2 = fig.add_subplot(gs[1, :])
labels_v = ["Long-duration permits", "Foreign-born under 30", "Migrant women in severe
poverty"]
values_v = [PCT_LONG_DURATION_PERMITS, PCT_FOREIGN_BORN_UNDER_30,
PCT_MIGRANT_WOMEN_SEVERE_POVERTY]
ax2.barh(labels_v, values_v)
ax2.set_xlim(0, 100)
ax2.set_xlabel("Percent (%)")
ax2.set_title("Selected migrant indicators")
for i, v in enumerate(values_v):
    ax2.text(v + 1, i, f"{v}%", va="center")
plt.tight_layout()
plt.show()
```

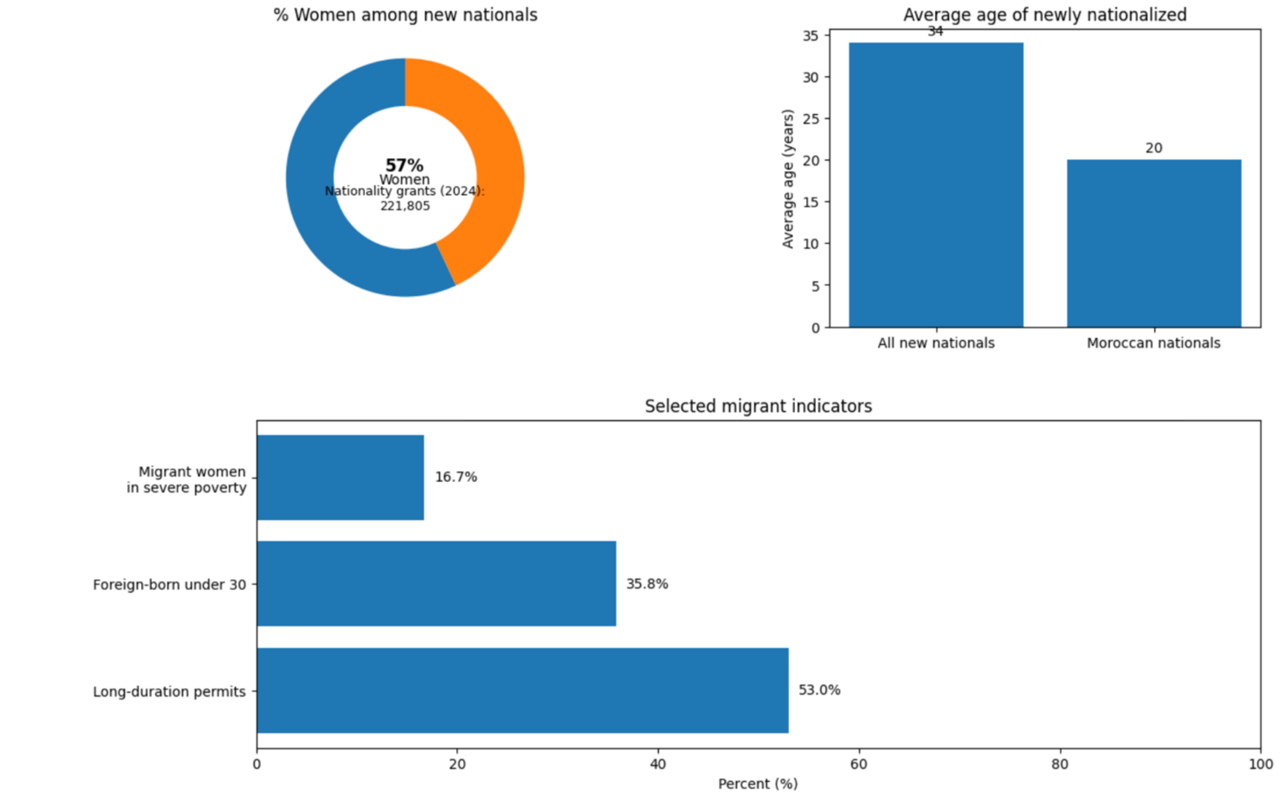


Figure 12 Examples of demographic indicators. The infographic combines gender balance among new nationals, age differences between groups, and key vulnerability measures to illustrate broader migration realities in Spain

The top-left donut emphasizes the gender composition of new nationals and embeds the absolute number of nationality grants as context, the top-right bar chart shows age differences between all new nationals and a specific group (here, Moroccan nationals), pointing to differing integration profiles, while the bottom panel surfaces structural vulnerabilities (permit stability, youth among the foreign-born, and severe poverty among migrant women). Together, these indicators offer a macro backdrop that complements the survey's micro-level experiences and perceptions.

Sports as integrator and challenge

In the BRISWA 2.0 questionnaire, the respondents overwhelmingly described sport as a tool for integration and respect across cultures. This optimism, however, sits in tension with the wider evidence. On the one hand, sport is repeatedly used as a platform for inclusion. An analysis of existing initiatives shows that many programs fall into five broad groups:

(i) sport as a tool for integration
(ii) community forums and dialogue
(iii) assistance to vulnerable individuals
(iv) awareness-raising actions
(v) education and training.

Among these, sport and football in particular emerges as a major entry point for efforts against racism.

On the other hand, football also appears frequently as a context of discrimination. Insights from analytical reports underline that, alongside public spaces, education, and online platforms, football is one of the arenas where racism is most visible.

```
import matplotlib.pyplot as plt
# --- Program groups (from WP2 analysis results) ---
program_groups = {
    "Sport as tool for integration": 1,
    "Community forums & dialogue": 1,
    "Support for vulnerable groups": 1,
    "Awareness-raising": 1,
    "Education & training": 1
}
# Each group = 1, we just want categorical display
# --- Contexts of discrimination (from Analytical Insights) ---
contexts = {
    "Public spaces": 57.1,
    "Education": 46.5,
    "Social media": 39.3,
    "Football": 7.1
}
fig, axes = plt.subplots(1, 2, figsize=(12, 5))
# Left: Program groups (equal counts, categorical bars)
axes[0].barh(list(program_groups.keys()), list(program_groups.values()),
             color=["#1a9850" if "Sport" in k else "#9ecae1" for k in
program_groups.keys()])
axes[0].set_title("Types of anti-racism programs")
axes[0].set_xticks([])
axes[0].set_xlim(0, 1.2)
for i, k in enumerate(program_groups.keys()):
    axes[0].text(0.5, i, k, ha="center", va="center", fontsize=9,
color="white", weight="bold")
# Right: Contexts of discrimination
axes[1].barh(list(contexts.keys()), list(contexts.values()), color="#fd8d3c")
axes[1].invert_yaxis()
axes[1].set_title("Contexts of discrimination")
axes[1].set_xlabel("Percent of 'Yes' respondents")
for i, (k, v) in enumerate(contexts.items()):
    axes[1].text(v + 1, i, f"{v:.1f}%", va="center")
plt.tight_layout()
plt.show()
```

On Figure 13, the left panel shows the five types of anti-racism programs identified in project analyses, with sport/football highlighted as a major vehicle for integration. The right panel shows the contexts of discrimination identified in analytical insights, where football also appears — not as the main arena, but as a recurring setting for racist incidents.Placed next to each other, these two perspectives mirror what we observed in the survey: enthusiasm about the potential of sport, but also recognition that its institutions and environments can reproduce exclusion. Sport therefore stands both as a means of integration and as a site of challenge.

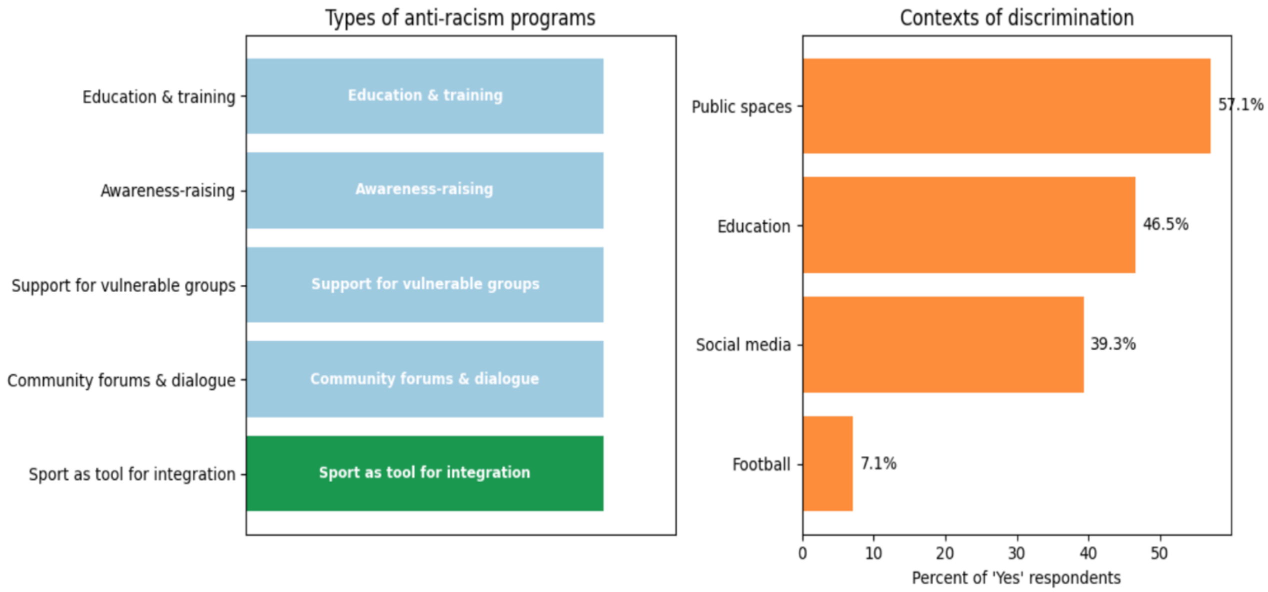


Figure 13 Sport as integration and challenge

Conclusions

Apart from their purpose in designing graphs for storytelling, the data that was used in the current deliverable reveal interesting insights. Firstly, they reveal an underlying optimism of the respondents with regards to optimism of institutions and programs to combat racism and discrimination, which can be attributed to the relatively young age of the majority of the respondents.

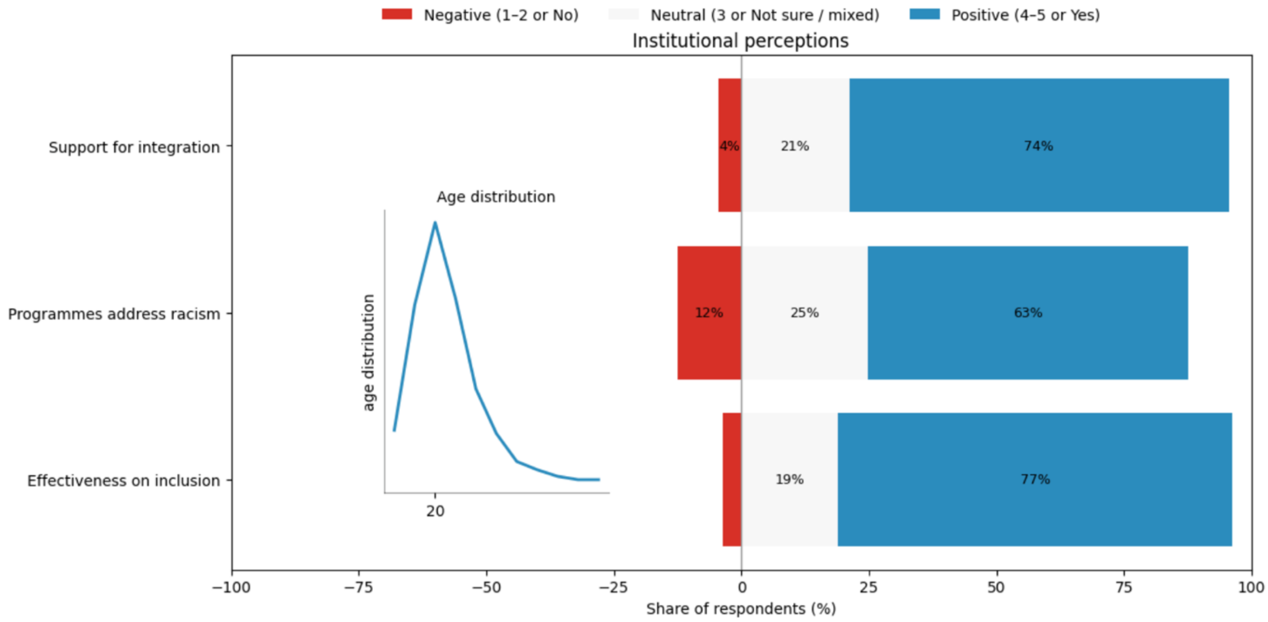


Figure 14 The underlying optimism of the responses to the BRISWA 2.0 survey indicated by the level of trust and relative age of the respondents.

Despite this optimism though, experiences of discrimination are not absent. In fact 1 in 5 respondents declared an experience of discrimination in various settings (that have the trust of the majority of the respondents), and only 3.6% of those that experienced discrimination asked or received assistance to deal with the ramifications of the incidence.

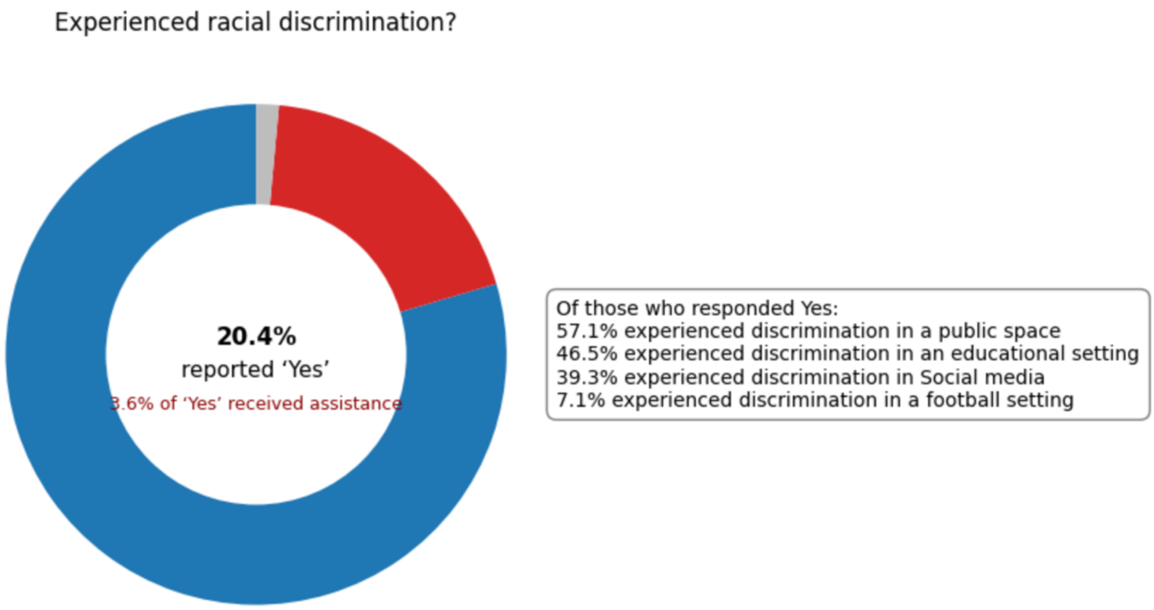


Figure 15 Discrimination insists despite the optimism

Especially football and sports, which are celebrated as a tool against racism, remain as sites of active discrimination experiences.

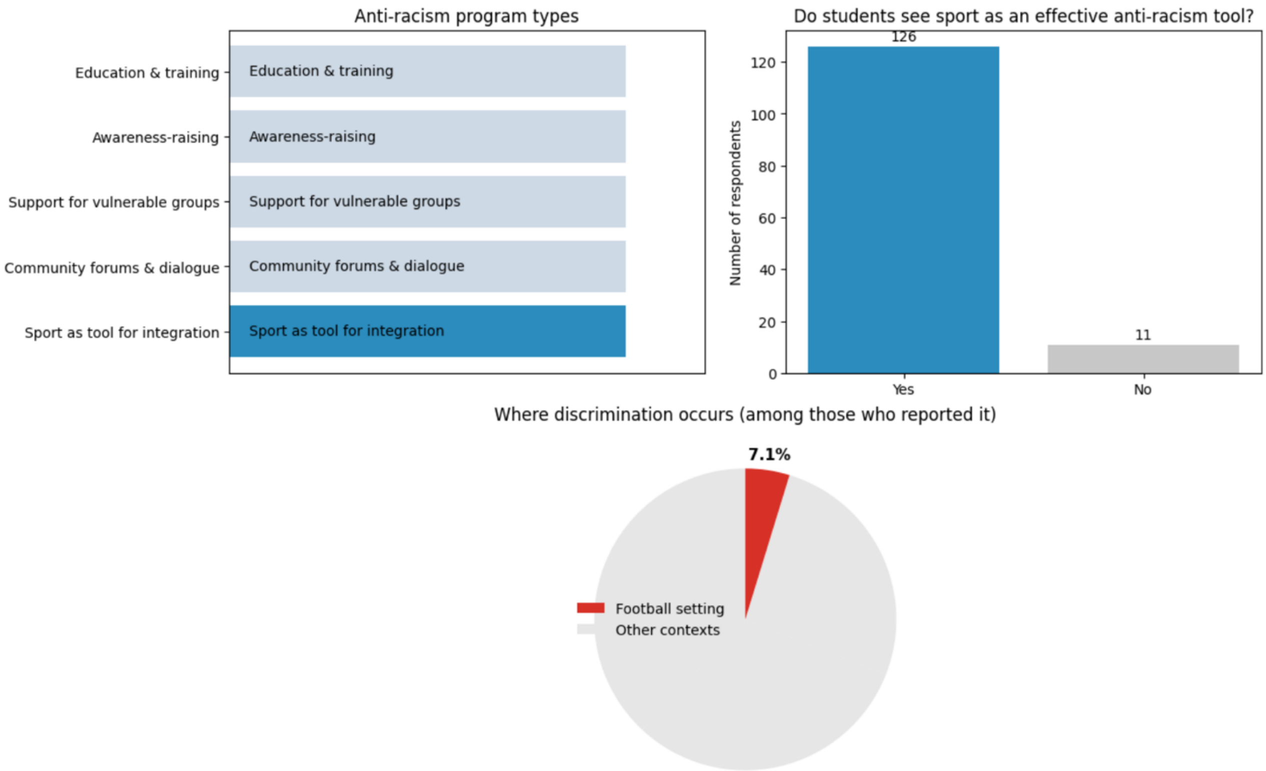


Figure 16 Sports are celebrated as an effective anti-racist tool, but remain an active site of discrimination

These insights reveal interesting implications. For public institutions and universities, optimism alone cannot be considered an effective policy. On the contrary, these organizations should establish mechanisms and frameworks to increase the trust of people and students, so that whenever

someone experiences discrimination, they could have the institutional support to report it and mitigate the personal ramifications.
In addition, sports programs have strong symbolic value, but they need as well mechanisms to transform this immense value into action that actively reduces racism and discrimination.
Finally, demographic realities cannot be independent of the discussion on how to combat racism.

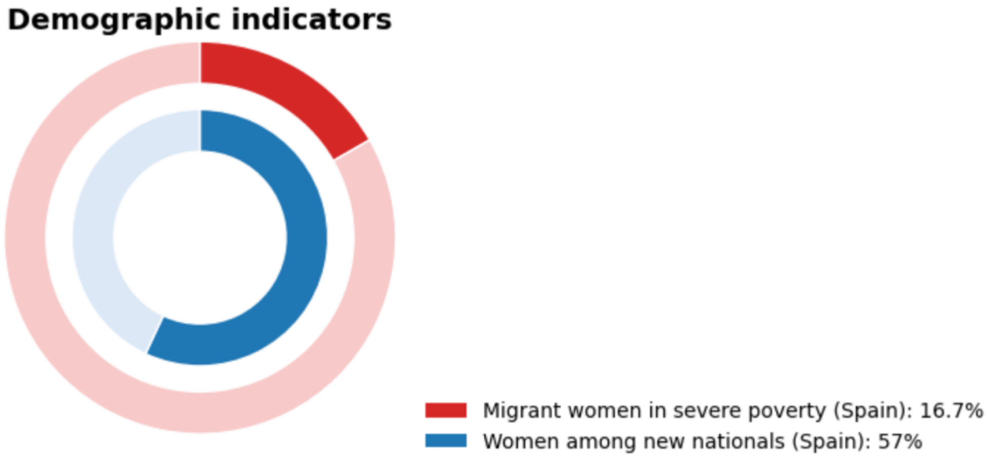


Figure 17 Demographic realities as context for the discussion on racism

The context must inform and guide policy design and program implementation, otherwise any type of decision is in danger of being ineffective from its inception.
Communicating sensitive social data requires particular attention to ethics, especially when the subjects of analysis belong to vulnerable or marginalized groups. Beyond accuracy and methodological transparency, ethical data storytelling demands respect for privacy, contextual integrity, and emotional impact. Graphs and narratives should avoid framing individuals or communities as problems to be solved, and should ensure that statistical representation does not erase personal experiences. When visualizing discrimination or exclusion, it is crucial to balance the need for evidence with the duty of care such as protecting anonymity, preventing harm, and fostering empathy and understanding.

References

Beheshti, A., Tabebordbar, A., & Benatallah, B. (2020, April). istory: Intelligent storytelling with social data. In Companion Proceedings of the Web Conference 2020 (pp. 253-256).

Bumblauskas, D., Nold, H., Bumblauskas, P. & Igou, A. (2017). Big data analytics: transforming data to action. Business Process Management Journal, Vol. 23 No. 3, pp. 703-720.

Boldosova, V., & Luoto, S. (2020). Storytelling, business analytics and big data interpretation: Literature review and theoretical propositions. Management Research Review, 43(2), 204-222.

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., & Shum, S. B.(2018). Driving data storytelling from learning design. International Conference on Learning Analytics and Knowledge, 8, 131-140. https://doi.org/10.1145/3170358.3170380

Erete, S., Ryou, E., Smith, G., Fassett, K. M., & Duda, S. (2016, February). Storytelling with data: Examining the use of data by non-profit organizations. In Proceedings of the 19th ACM conference on Computer-Supported cooperative work & social computing (pp. 1273-1283).

Gershon, N. & Page, W. (2001). What storytelling can do for information visualization. Communications of the Acm, Vol. 44 No. 8, pp. 31-37.

Hullman, J. & Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. IEEE Transactions on Visualization and Computer Graphics, 17(12):2231–2240.

Islam, M. S., Laskar, M. T. R., Parvez, M. R., Hoque, E., & Joty, S. (2024). DataNarrative: Automated data-driven storytelling with visualizations and texts. arXiv preprint arXiv:2408.05346.

Kelliher, A., & Slaney, M. (2012). Tell me a story. IEEE multimedia, 19(1), 4-6. https://doi.org/10.1109/MMUL.2012.13

Knaflic, C. N. (2015). Storytelling with data. Wiley.

Lotfi, F., Beheshti, A., Farhood, H., Pooshideh, M., Jamzad, M., & Beigy, H. (2023). Storytelling with image data: a systematic review and comparative analysis of methods and tools. Algorithms, 16(3), 135.

Lund, B. D. (2022). The Art of (Data) Storytelling. The International Journal of Information, Diversity, & Inclusion, 6(1/2), 31-41.

Matei, S. A., & Hunter, L. (2021). Data storytelling is not storytelling with data: A framework for storytelling in science communication and data journalism. The Information Society, 37(5), 312-322.

Otten, J., Cheng, K., & Drewnowski, A. (2015). Infographics and public policy: Using data visualization to convey complex information. Health Affairs, 34(11):1901–1907.

Shi, D., Xu, X., Sun, F., Shi, Y., & Cao, N. (2020). Calliope: Automatic visual data story generation from a spreadsheet. IEEE Transactions on Visualization and Computer Graphics, 27(2), 453-463.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Voida E., Harmon, E. & Al-Ani, B. (2011). Homebrew Databases: Complexities of Everyday Information Management in Nonprofit Organizations. In Proc. of CHI (CHI ’11). ACM, New York, NY, USA, 915–924. DOI: http://dx.doi.org/10.1145/1978942.1979078

Yang, C. (2013). Telling tales at work. Business Communication Quarterly, Vol. 76 No. 2, pp. 132-154.