

# TWD – Homework 3

## Porównanie dobrych i złych wykresów

Sabina Sidarovich, Jan Krężel

### 1 Wstęp

#### 1.1 Cel

Celem eksperymentów było porównanie czytelności wykresów stworzonych w zgodzie z "dobrymi praktykami" z czytelnością wykresów wykonanych "niepoprawnie".

#### 1.2 Krótki opis eksperymentów

Dwa przeprowadzone eksperymenty miały podobne założenie. Ten sam zestaw danych (inny dla eksperymentów) przedstawiliśmy na dwa różne sposoby, starając się aby jedyną różnicą między wizualizacjami był jej styl. To znaczy w obu przypadkach dane pokazywane były w tej samej kolejności oraz z takimi samymi wskazówkami wizualnymi (siatka, odstęp między przedziałkami, kolor itd.)

#### 1.3 Metodologia

Do zgromadzenia danych o czytelności wykresów stworzyliśmy ankietę, która losowała jeden z dwóch zestawów pytań. Oba zestawy pełniły rolę grupy kontrolnej dla jednego z eksperymentów oraz grupy eksperymentalnej dla drugiego. Innymi słowy nikt nie widział dwóch wizualizacji tego samego zestawu danych i dla każdego ankietowanego część pytań dotyczyła wykresu poprawnego, a część niepoprawnego.

Nie chcieliśmy, aby ktokolwiek widział ten sam zestaw danych dwukrotnie,

ponieważ zobaczenie jednej z nich mogłoby wpłynąć na odpowiedzi dotyczące drugiej.

Zestawy były losowane, dlatego ilość odpowiedzi do obu z nich nieznacząco się różni (22 i 24). Przy analizie wyników zajmujemy się jedynie rozkładem wszystkich odpowiedzi.

## 2 Eksperyment 1

### 2.1 Opis

Celem eksperymentu było porównanie czytelności wykresu słupkowego z czytelnością wykresu kołowego bez wskazówek wizualnych<sup>1</sup>. Do obu wizualizacji zadaliśmy jedno i to samo pytanie

*Który kolor reprezentuje największą wartość? (Jeśli kilka zaznacz wszystkie)*

### 2.2 Dane

W poniższej tabelce umieszczone są wizualizowane dane.

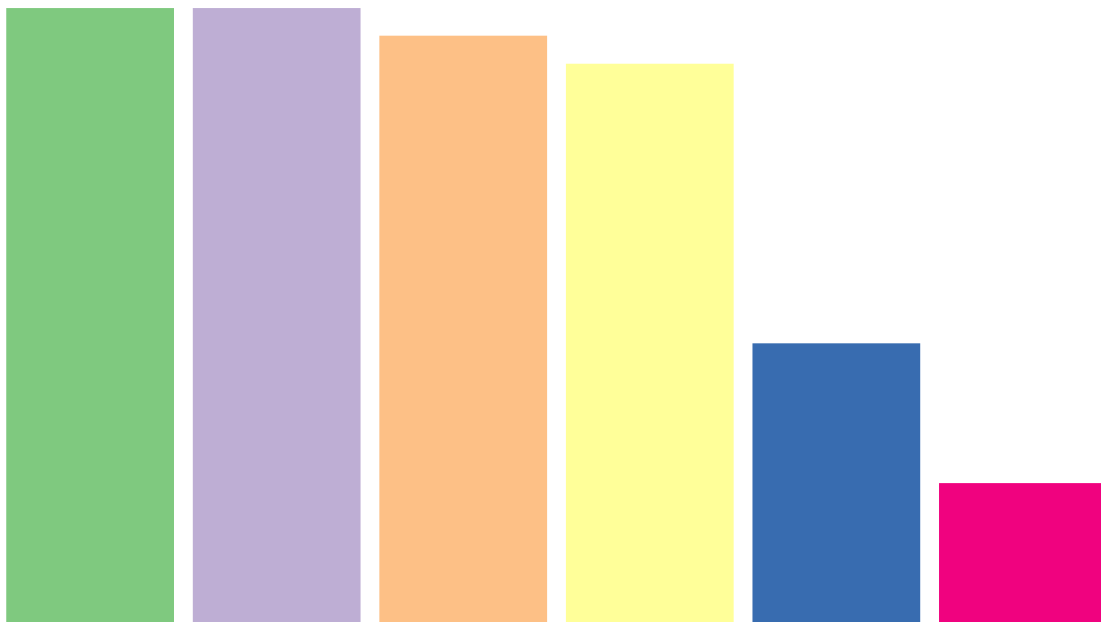
Name	Value
Green	22
Purple	22
Orange	21
Yellow	20
Blue	10
Pink	5

Tabela 1: Dane do eksperymentu 1

---

<sup>1</sup>dlatego w żadnej z wizualizacji nie ma osi ani przedziałek

## 2.3 Wizualizacja 1



Rysunek 1: Wykres słupkowy pierwszego zestawu danych

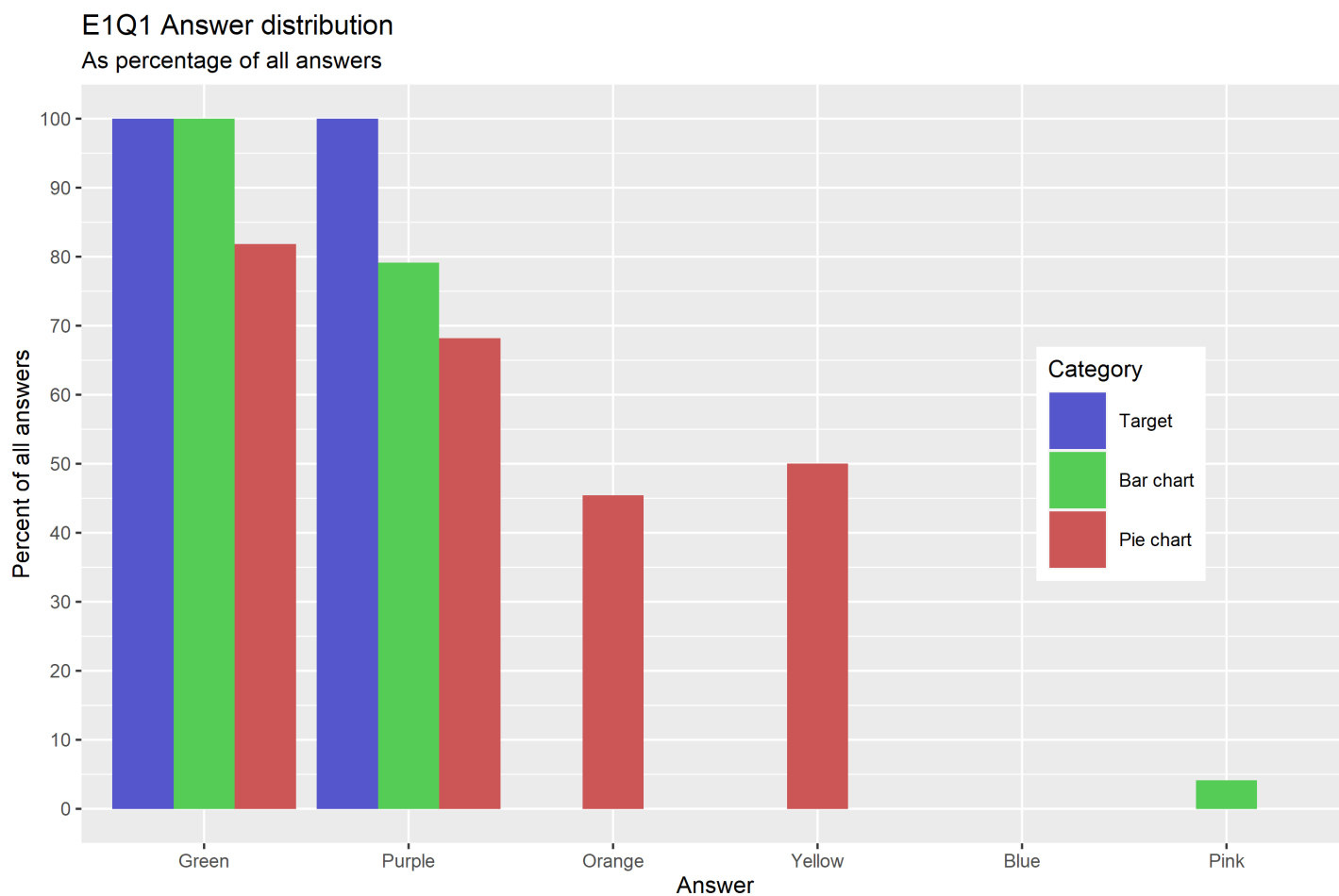
## 2.4 Wizualizacja 2



Rysunek 2: Wykres kołowy pierwszego zestawu danych

## 2.5 Wyniki

### 2.5.1 Wizualizacja wyników



Rysunek 3: Wizualizacja wyników pierwszego eksperymentu

Kategoria *Target* oznacza docelowy rozkład odpowiedzi.

### 2.5.2 Omówienie wyników

Z powyższego wykresu możemy wywnioskować, że wykres słupkowy dużo lepiej wizualizował nasze dane. Rozkład odpowiedzi niemal idealnie odzwier-

ciędlu ten docelowy. Pojedyncza odpowiedź "Pink" mogła być żartem jednego z ankietowanych lub skutkiem nieuważnego doboru kolorów. Osoba, która zaznaczyła "Pink" nie zaznaczyła "Purple", co może oznaczać, że odwrotnie rozróżniła te dwa kolory.

Rozkład odpowiedzi dotyczących wykresu kołowego wyraźnie obrazuje problemy związane z tym typem wykresu. Żaden kolor nie został wybrany przez wszystkich ankietowanych, a kolory, które w przypadku wykresu słupkowego nigdy nie zostały wybrane, tutaj wybierane były niemal na równi z poprawnymi.

Wynik tego testu jasno wskazuje, że wykresy kołowe nie tylko są mniej czytelne od wykresów słupkowych, ale także wprowadzają odbiorcę w błąd.

## 3 Eksperyment 2

### 3.1 Opis

Celem eksperymentu było porównanie czytelności wykresu słupkowego, w którym słupki odpowiadające danym kategoriom postawione są obok siebie z czytelnością wykresu słupkowego, w którym te słupki ułożone są na stosie. Do obu wizualizacji zadaliśmy trzy pytania

1. *Ilu kandydatów z grupy A przyjął wydział 1?*
2. *Ilu kandydatów z grupy A przyjął wydział 2?*
3. *Który wydział przyjął większą liczbę kandydatów z grupy A?*  
(3, 4, Oba tyle samo)

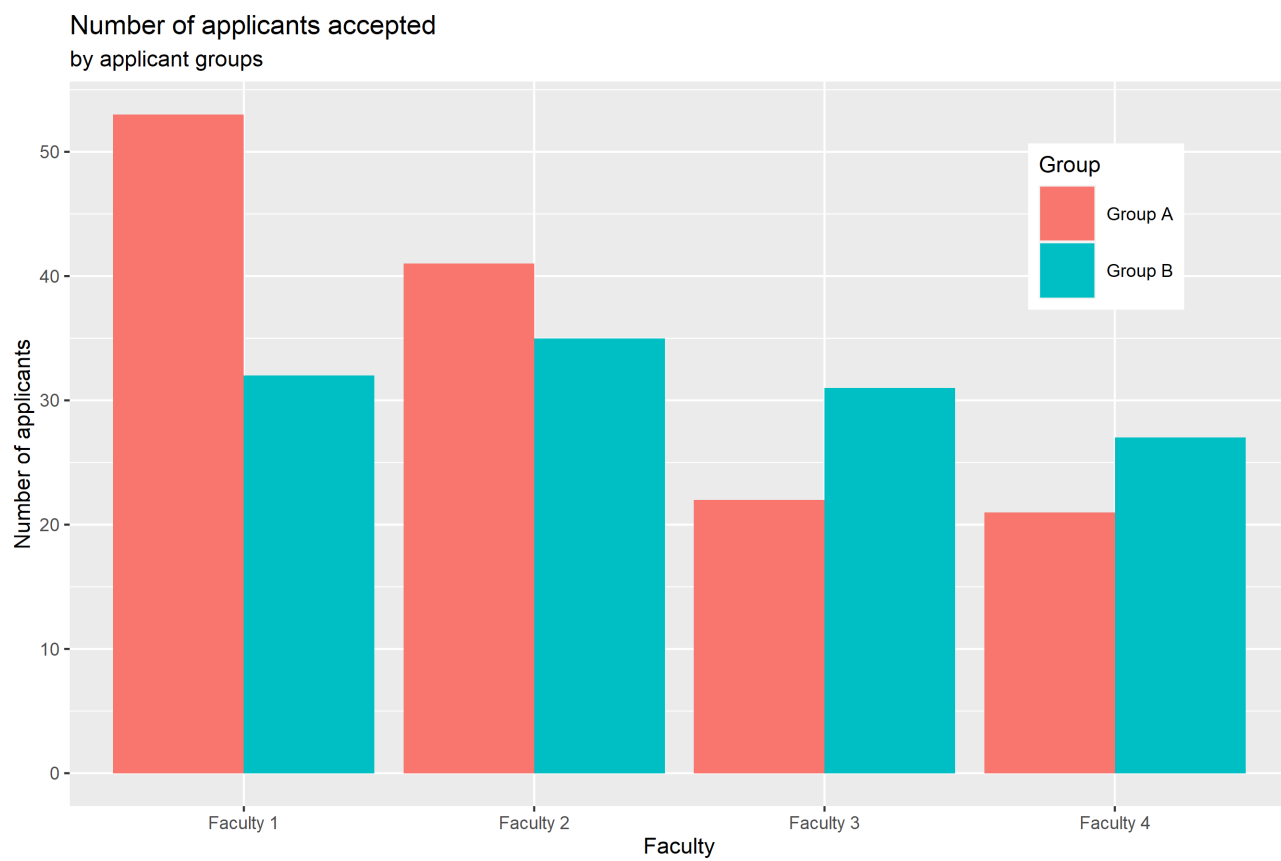
### 3.2 Dane

W poniższej tabelce umieszczone są wizualizowane dane.

Faculty	Group	Acceptees
Faculty 1	Group A	53
Faculty 1	Group B	32
Faculty 2	Group A	41
Faculty 2	Group B	35
Faculty 3	Group A	22
Faculty 3	Group B	31
Faculty 4	Group A	21
Faculty 4	Group B	27

Tabela 2: Dane do eksperymentu 2

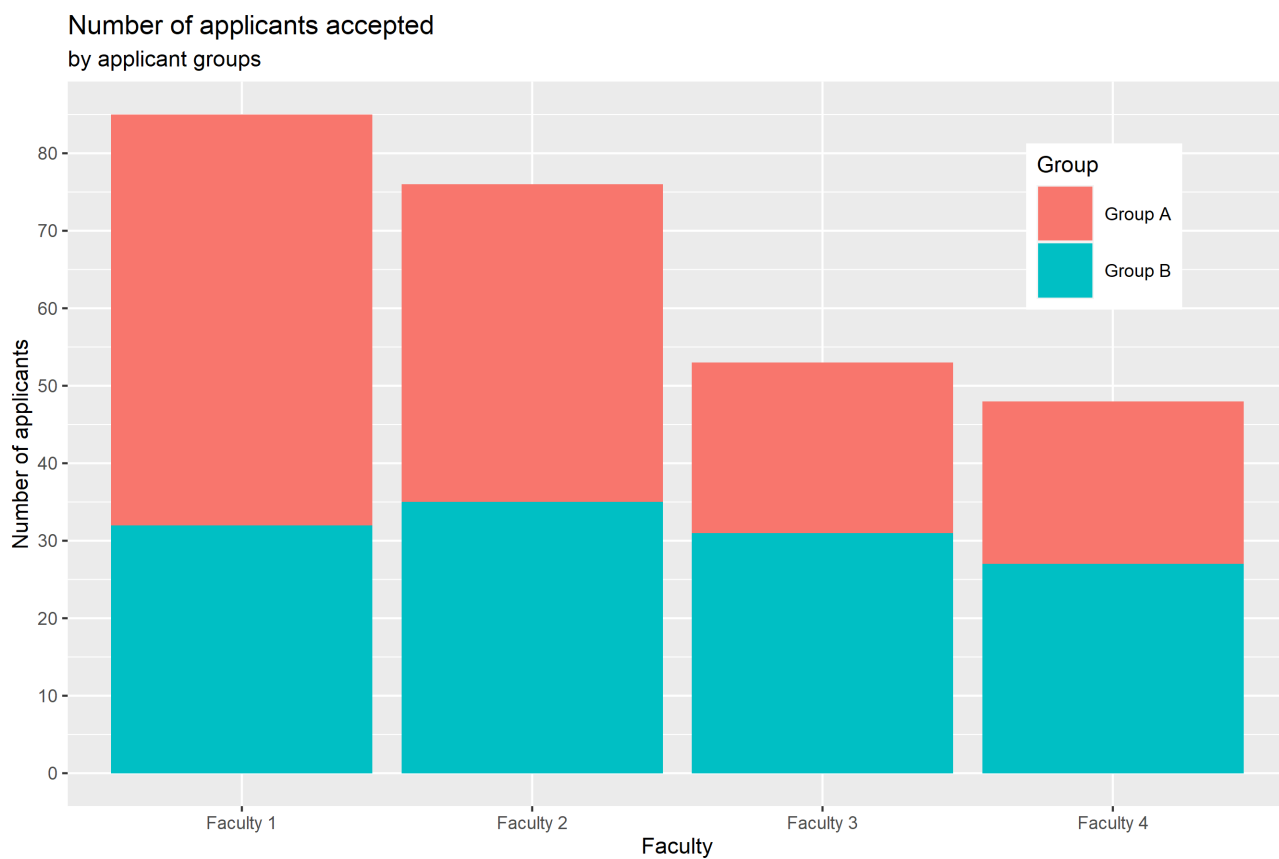
### 3.3 Wizualizacja 1 (Dodge)



Rysunek 4: Wykres słupkowy z kolumnami obok siebie



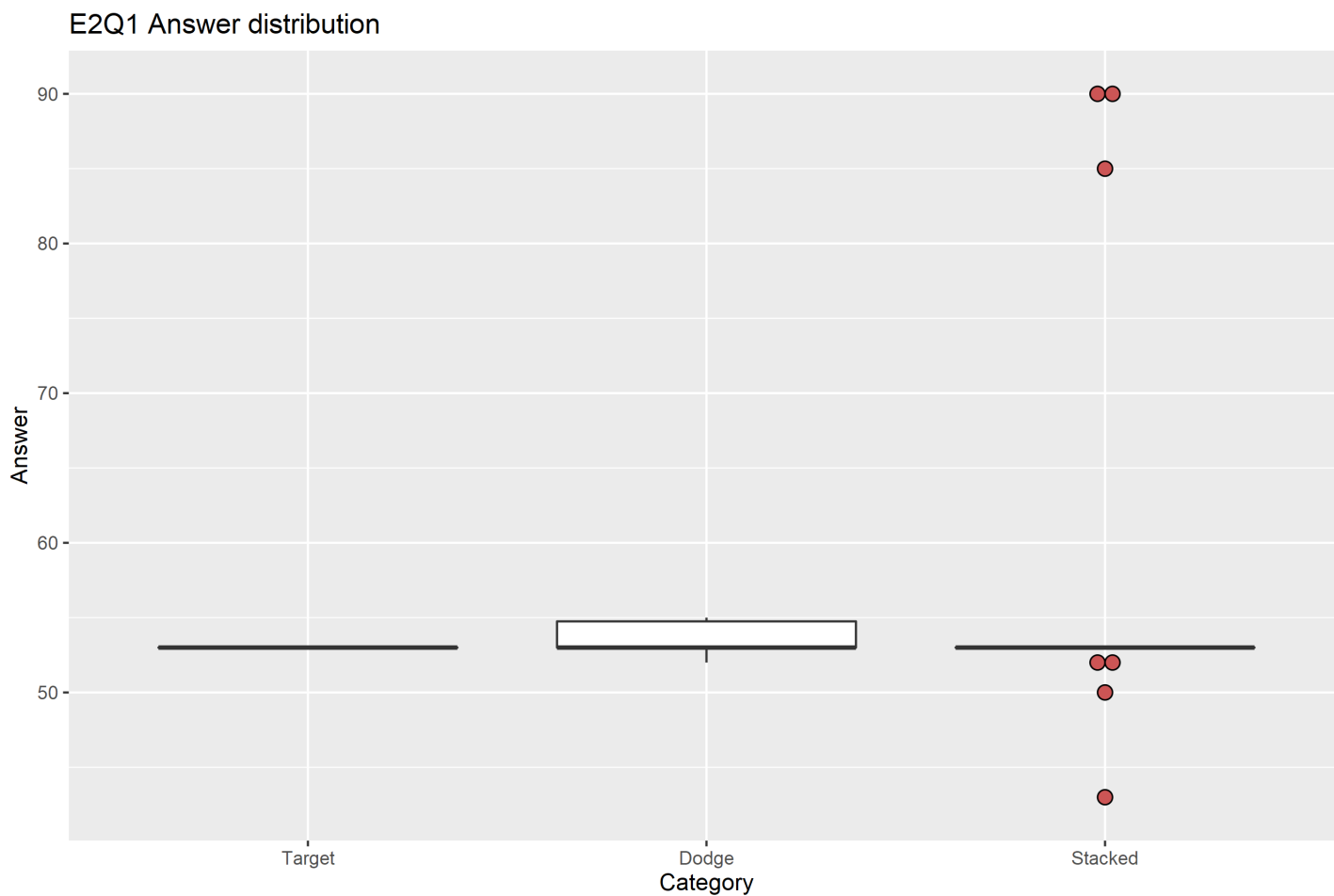
### 3.4 Wizualizacja 2 (Stacked)



Rysunek 5: Wykres słupkowy z kolumnami na stosie

## 3.5 Wyniki – Pytanie 1

### 3.5.1 Wizualizacja wyników



Rysunek 6: Wizualizacja wyników pierwszego pytania drugiego eksperymentu

Kategoria *Target* spełnia podobną rolę jak w pierwszym eksperymencie. Jest ona punktem odniesienia, do którego porównujemy oba zestawy odpowiedzi.

### 3.5.2 Omówienie wyników

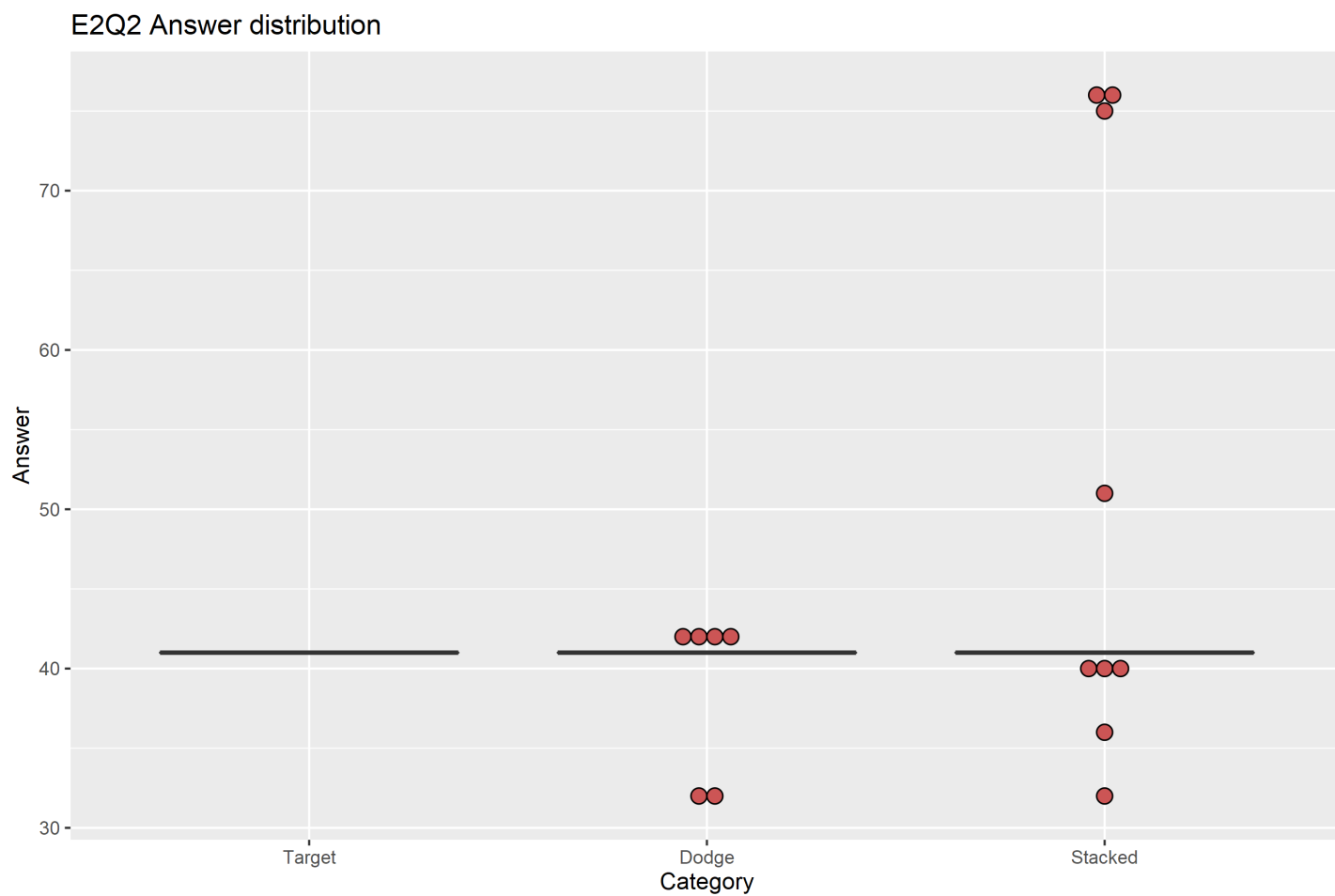
Na powyższym wykresie widać, że mediana odpowiedzi dla obu wizualizacji jest bardzo bliska docelowej.

Wiele odpowiedzi dotyczących wykresu *Dodge* była równa 55. Może to wynikać z niedokładnej analizy przedziałek na pionowej osi. Faktycznie czerwony słupek wydaje się kończyć mniej więcej w połowie między przedziałkami, zaraz po 50, co ankietowani mogli odczytywać jako 55. Być może gdyby oś kończyła się na wartości 60 ten błąd nie pojawiałby się aż tak często.

Dla wykresu *Stacked* mamy dużo więcej odpowiedzi odstających. W szczególności dwie odpowiedzi równe 90 wskazują na największy problem z tą wizualizacją – brak jednoznaczności w tym czy czerwony słupek zaczyna się w zerze czy zaraz nad niebieskim. Odpowiedź 85 to prawdopodobnie połączenie tej niejednoznaczności ze złym odczytaniem przedziałek (podobnie jak w przypadku odpowiedzi 55). Błędne odpowiedzi poniżej mediany to prawdopodobnie błędy wynikające z potrzeby odejmowania.

## 3.6 Wyniki – Pytanie 2

### 3.6.1 Wizualizacja wyników



Rysunek 7: Wizualizacja wyników drugiego pytania drugiego eksperymentu

### 3.6.2 Omówienie wyników

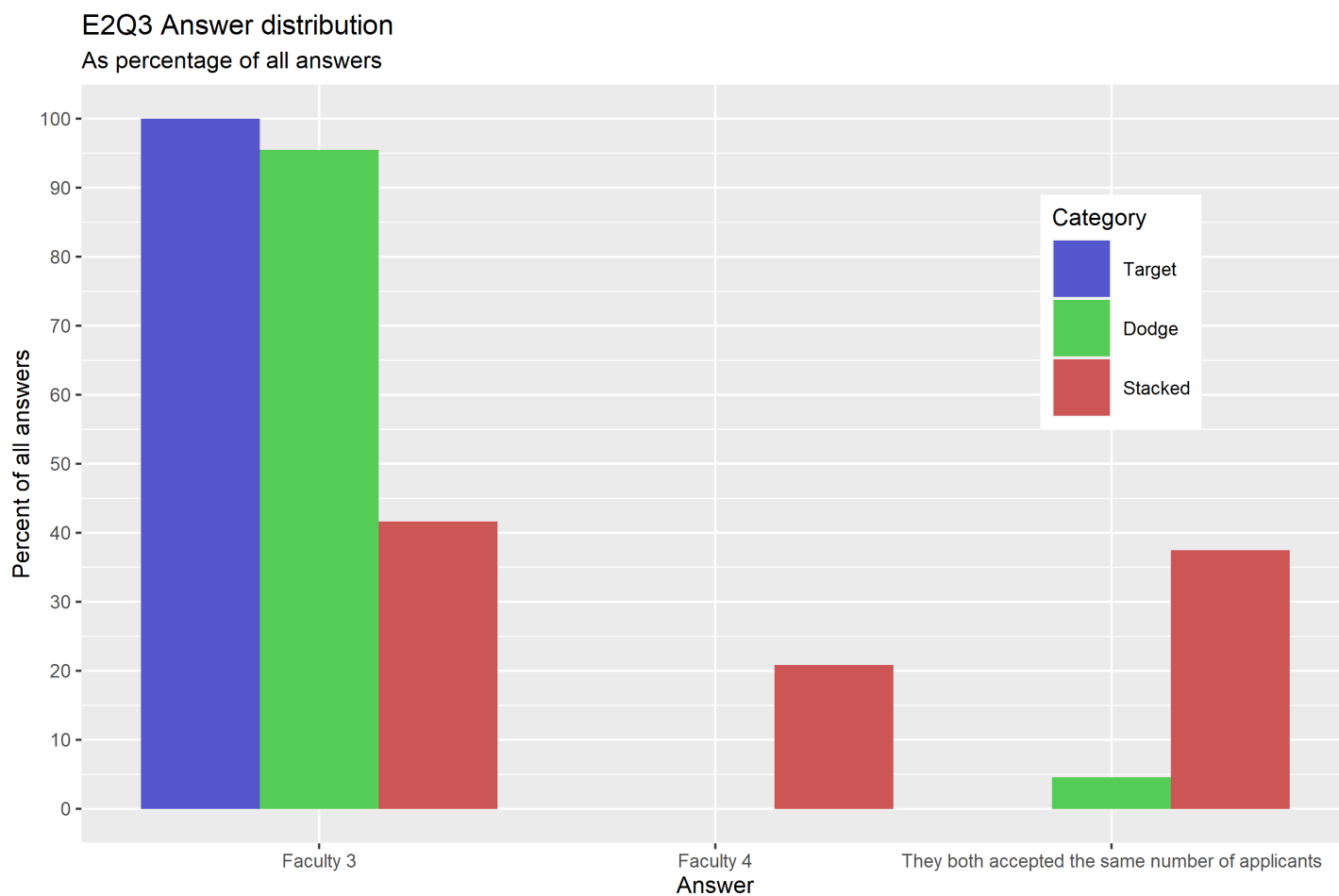
Podobnie jak w przypadku pierwszego pytania mediana odpowiedzi znów jest bardzo zbliżona do docelowej.

Dla wykresu *Dodge* mamy 6 błędnych odpowiedzi, które można podzielić na dwie grupy. Pierwsza z nich to odpowiedzi równe 42 (choć jest to poprawna odpowiedź na inne bardzo ważne pytanie). Ten błąd najprawdopodobniej wynika po prostu ze złego odczytania wysokości słupka. Druga grupa to dwie odpowiedzi równe 32. Mogą one wynikać z niedokładnego przeczytania pytania, ponieważ jest to liczba kandydatów z **grupy B** zaakceptowanych przez **Wydział 1** (czyli ktoś mógł zamienić grupę a nie wydział z pierwszego pytania).

Dla wykresu *Stacked* trzy najbardziej oddalone odpowiedzi znów wynikają z niejednoznaczności początku czerwonego słupka. Tutaj również mamy jedną odpowiedź 32. Pozostałe błędne odpowiedzi to mieszanka błędów w odejmowaniu i trudności w odczytaniu dwóch wartości zamiast jednej.

## 3.7 Wyniki – Pytanie 3

### 3.7.1 Wizualizacja wyników



Rysunek 8: Wizualizacja wyników trzeciego pytania drugiego eksperymentu

### 3.7.2 Omówienie wyników

Z powyższego wykresu od razu widać, że to pytanie sprawiło najwięcej trudności grupie czytającej z wykresu *Stacked*.

Odpowiedzi dotyczące wykresu *Dodge* są niemal perfekcyjne. Tylko jedna osoba odpowiedziała na to pytanie źle.

Osoby odpowiadające na podstawie wykresu *Stacked* wpadały w pułapkę iluzji optycznej i twierdziły, że oba wydziały przyjęły taką samą ilość kandydatów lub nawet, że więcej z nich przyjął wydział 4. Co ciekawe, analizując poprzednie dwa pytania, mogłoby się wydawać, że odczytywanie dwóch wartości i odjęcie ich od siebie nie sprawia dużego kłopotu, a wykonanie tych czynności dla obu wydziałów i porównanie różnic powinno dać poprawną odpowiedź. Być może ankietowani czuli się na tyle pewnie porównując "na oko" dwa słupki, które na wykresie wizualnie są obok siebie, że nie czuli potrzeby weryfikacji swojej odpowiedzi.

## 4 Podsumowanie

### 4.1 Podsumowanie wyników

Poprawne wykresy były odczytywane bardzo dokładnie, a jakiegokolwiek błędne odpowiedzi najczęściej wynikały z innych czynników (zły dobór kolorów, źle przeczytane pytanie, źle dobrane przedziały).

Niepoprawne wykresy notorycznie były odczytywane błędnie. Tutaj błędy dużo częściej miały swoje źródło w niejednoznaczności wizualizacji, iluzjach optycznych, potrzeby wykonania większej ilości odczytów czy potrzeby odejmowania niektórych z nich.

Analizując wszystkie wyniki możemy dojść do wniosku, że "dobre praktyki" faktycznie znacząco poprawiają czytelność wykresów i powinno się je zawsze stosować. Przedstawianie danych w sposób, który nie przestrzega tych praktyk jest jednoznacznie gorsze, a użycie tego typu wykresów może oznaczać niekompetencję lub chęć zmylenia odbiorcy bez potrzeby manipulacji danymi.

## 4.2 Wiadome wady eksperymentów

Prawdopodobnie największą wadą eksperymentów jest nieodpowiedzialny dobór kolorów. Korzystaliśmy ze standardowych palet wbudowanych w pakiety `ggplot2` i `RColorBrewer` jednak w pierwszym eksperymencie miało to realny wpływ na wynik, gdy jeden z ankietowanych prawdopodobnie pomylił kolor fioletowy z różowym. W drugim eksperymencie potencjalnie mogły pojawić się problemy dla osób z daltonizmem. Nie jest to problem, który zawsze da się rozwiązać, jednak w tym przypadku korzystaliśmy tylko z dwóch kolorów, które łatwo można było lepiej rozróżnić.

Mogliśmy także uwzględnić faktyczną grupę kontrolną, której zadaniem byłoby odczytywanie surowych danych z tabeli. Jednak odpowiedzi takiej grupy najprawdopodobniej całkowicie pokrywałyby się z kategorią *Target*, a wymagałoby to stworzenia większej ilości zestawów pytań i podzielenie ankietowanych na jeszcze mniejsze grupki.

Kolejnym problem, który potencjalnie mógł mieć wpływ na odpowiedzi była ich ustalona kolejność. W szczególności osoby odpowiadające na pytanie dotyczące wykresu *Stacked* dostawały je w tej samej kolejności. Po odpowiedzi na dwa, potencjalnie nieprzyjemne, pytania, ankietowani mogli nie przykładąć równie wysokiej uwagi do trzeciego, najtrudniejszego, pytania. Pomieszanie kolejności pytań mogłoby zmienić rozkład odpowiedzi na korzyść tej wizualizacji.