

# Braki odpowiedzi w wybranych badaniach reprezentacyjnych gospodarstw domowych

BAEL, BBGD, EU-SILC, ICT

Sebastian Kuzara & Maciej Beręsewicz (Urząd Statystyczny w Poznaniu)

# Agenda

1. Braki odpowiedzi w badaniach reprezentacyjnych - skutki, definicja, rodzaje
2. Przyczyny braków odpowiedzi
3. Schematy losowania i organizacji badań reprezentacyjnych - porównanie
4. Wskaźnik realizacji w badaniach reprezentacyjnych (ogólnie i w przekrojach)
5. Przykłady - prezentacja na wykresach danych dotyczących braków odpowiedzi/wskaźników realizacji w wybranych przekrojach
6. Podsumowanie i rekomendacje

# Błędy w badaniach reprezentacyjnych

- Błędy losowe - związane z precyzją wyników w próbie
- Błędy nielosowe:
  - błędy operatu
  - pomiaru (niezrozumiała instrukcja)
  - odpowiedzi (niechęć udzielania prawdziwych odpowiedzi)
  - błędy spowodowane brakami odpowiedzi

# Skutki braków odpowiedzi

- Zmniejszenie liczebności próby, co wpływa na zwiększenie wariancji estymatorów
- Braki odpowiedzi, jeżeli występują w dużej liczbie, mogą negatywnie wpływać na jakość wyników
- Struktura próby zbadanej różni się od struktury próby wylosowanej pod względem cech społeczno-demograficznych
- Zwiększenie obciążenia estymatora

# Definicja

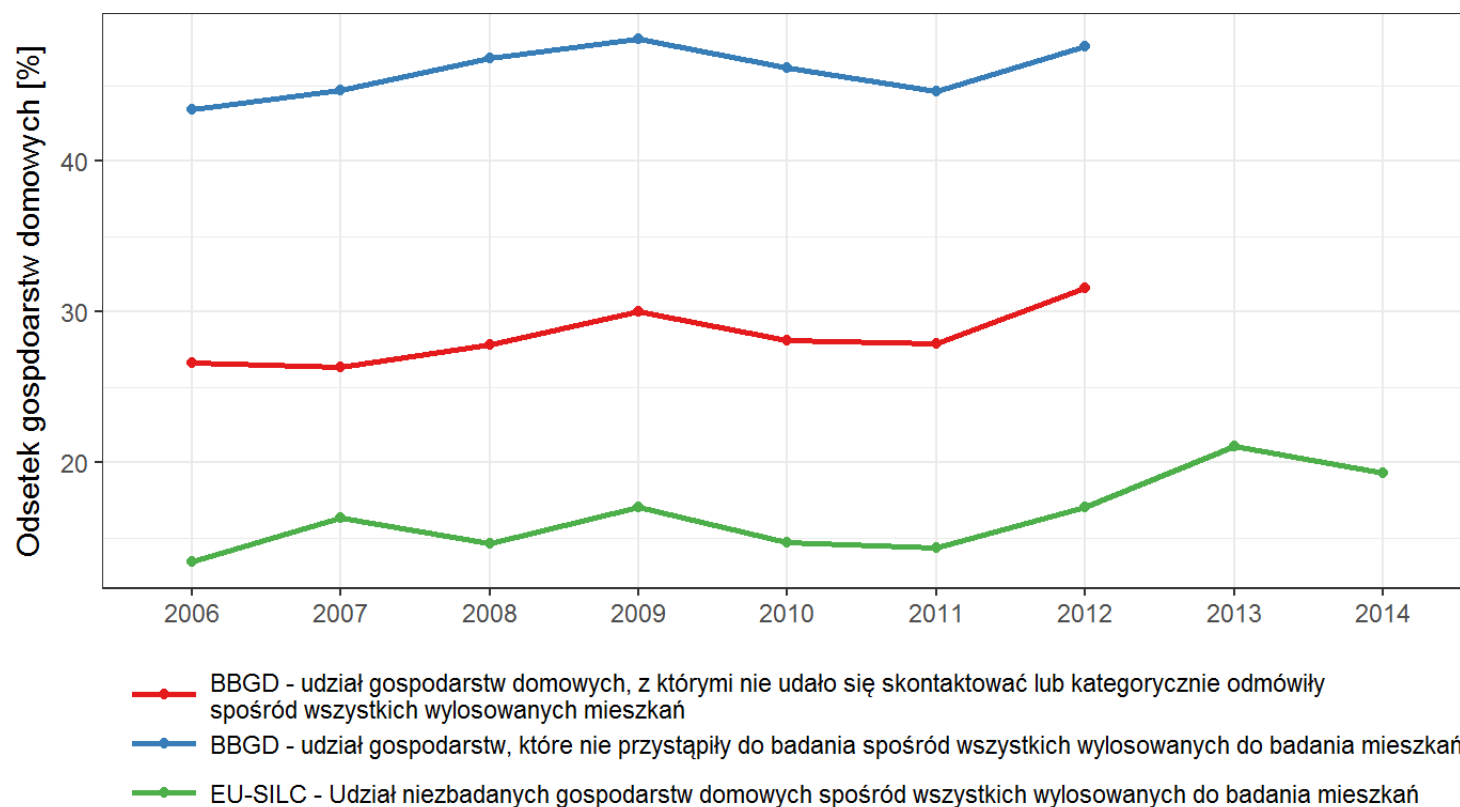
Brak odpowiedzi - sytuacja, w której, pomimo próby kontaktu ze strony ankietera, nie jest możliwe uzyskanie od respondenta informacji na pytania zawarte w ankiecie niezależnie od przyczyny, pod warunkiem że wylosowane do próby mieszkanie istnieje, pełni funkcje mieszkaniowe, jest zamieszkałe przez respondentów spełniających założone w badaniu kryteria lub istnieje takie podejrzenie\*.

\*Definicja własna na podstawie raportów z badań BGD, AEL, EU-SILC, ICT oraz Standard Definitions...

# Rodzaje braków odpowiedzi

- Całkowity brak odpowiedzi
- Częściowy brak odpowiedzi

# Rodzaje braków odpowiedzi w BBGD i EU-SILC



# Z czego wynikają braki odpowiedzi?

- Przyczyny obiektywne:
  - Utrudniony kontakt np. z powodu choroby
  - Czasowa nieobecność
  - Brak dostępu do mieszkania
  - Rozpad gospodarstwa domowego
  - Nikogo nie zastano w mieszkaniu
- Przyczyny subiektywne:
  - Odmowa udzielenia odpowiedzi
  - Niechęć
  - Brak czasu

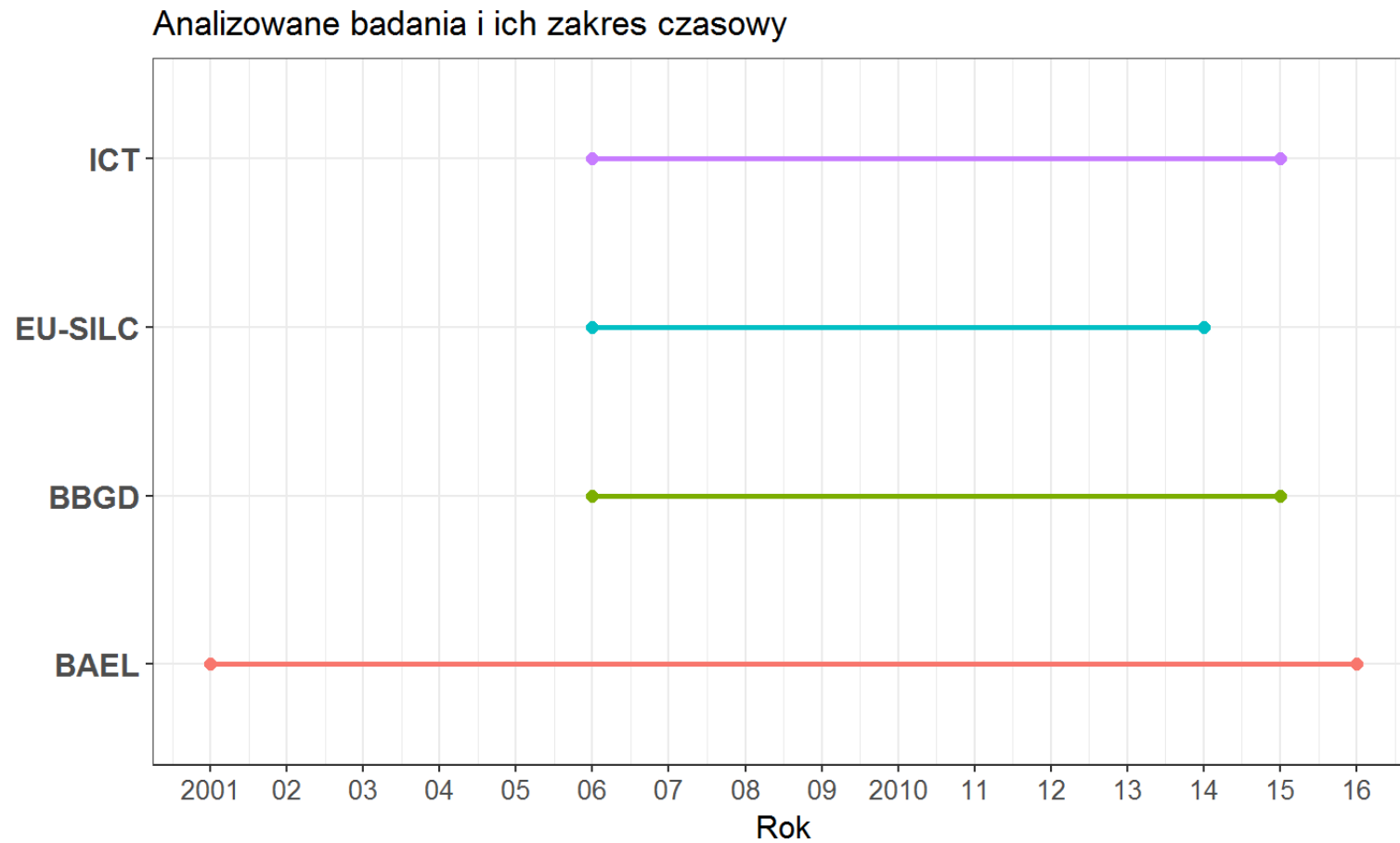


# Przyczyny braków odpowiedzi w BBGD i EU-SILC

# Kategorie przyczyn niepodjęcia udziału w badaniu

EU.SILC	BBGD
odmowa	odmowa
utrudnony kontakt	utrudnony kontakt
czasowa nieobecność	czasowa nieobecność
inne	inne
	nikogo nie zastano
	brak dostępu do mieszkania
	rozpad gospodarstwa domowego

# Badania



# Schematy losowania - porównanie

Wyszczególnienie	BBGD	EU.SILC	BAEL	ICT
metoda losowania	losowanie dwustopniowe z warstwaniem na pierwszym stopniu	losowanie dwustopniowe z warstwaniem na pierwszym stopniu	losowanie dwustopniowe z warstwaniem na pierwszym stopniu	losowanie dwustopniowe z warstwaniem na pierwszym stopniu
jednostka pierwszego stopnia	<p>rejony statystyczne* według spisów przygotowanych na potrzeby NSP z późniejszymi rocznymi aktualizacjami**</p> <p>* rejon statystyczny w miastach liczy około 250 mieszkań a na terenach wiejskich około 150</p> <p>** łącznie około 30 tys. JPS</p>	obwody spisowe	<p>miasta: rejony statystyczne</p> <p>tereny wiejskie: obwody spisowe</p> <p>obwody spisowe&gt;rejony statystyczne</p> <p>• występują (rzadko) sytuacje łączenia kilku rejonów/obwodów w jeden JPS</p>	rejony statystyczne według danych utworzonych na rzecz NSP miasta: 250 mieszkań tereny wiejskie 150 mieszkań
jednostka drugiego stopnia	zamieszkałe mieszkania według wykazów sprządzonych przez urzędy statystyczne	zamieszkałe mieszkania z listy sporządzonej na podstawie TERYT	mieszkania losowane z listy stworzonej na podstawie Krajowego Rejestru Podziału Terytorialnego Kraju (TERYT)	mieszkania

# Warstwowanie JPS

BBGD	EU.SILC	BAEL	ICT
1. województwa 2. 6 klas miejscowości 3. w Warszawie warstwami są dzielnice (łączone) 4. duże miasta tworzą samodzielne warstwy 5. małe miasta i tereny wiejskie według NTS3 • 191 warstw w tym 58 wiejskich • warstwy terenów wiejskich tworzone z połączenia terenów wiejskich z kilku powiatów	1. województwa 1. klasy miejscowości 2. 5 największych miast - dzielnice 3. duże miasta - samodzielne warstwy 4. tereny wiejskie - gminy wiejskie w podregionie (lub w kilku powiatach) 5. tereny miejskie według wielkości miast • 211 warstw	1. województwa 1. klasy miejscowości • od 3 do 7 warstw wewnątrz- wojewódzkich • brak sztywnego kryterium podziału na warstwy wewnątrz- wojewódzkie	1. województwa 1. klasy miejscowości 2. aglomeracje miejskie - osobne warstwy • 96 warstw

# Losowanie JPS (1)

BBGD	EU.SILC	BAEL	ICT
1. liczba losowanych JPS proporcjonalna do liczby mieszkań w warstwach	1. liczba losowanych JPS z pojedynczej warstwy równa się liczbie mieszkań do wylosowania z warstwy podzielonej przez liczbę mieszkań dla konkretnej klasy miejscowości	1. Metoda ekspercka ustalania liczby JPS losowanych z warstwy	1. liczba losowanych JPS w warstwie proporcjonalna do liczby mieszkań w warstwie
<p>1. Brak informacji na temat schematu losowania JPS*</p> <p>* losowanie dodatkowo pewnej liczby JPS wiejskich z puli JPS terenów wiejskich</p>	<p>1. losowanie za pomocą schematu Hartley'a-Rao:</p> <ul style="list-style-type: none"> <li>• posortowanie losowo JPS w warstwie</li> <li>• utworzenie zmiennej ze skumulowaną sumą wartości zmiennej, która ustala prawdopodobieństwo wylosowania JPS</li> <li>• systematyczny wybór co <math>\text{sum}(x)/n</math> jednostki według zmiennej ze skumulowaną sumą</li> </ul>	1. prawdopodobieństwo wylosowania JPS z warstwy proporcjonalne do liczby mieszkań w JPS	1. Losowanie wg schematu Hartley'a-Rao (patrz EU-SILC)

# Losowanie JPS (2)

BBGD	EU.SILC	BAEL	ICT
1. Próba dla każdego roku jest dzielona na dwie równe podpróbki. Każda z nich oprócz badanego roku bazowego obejmuje swoim zakresem rok poprzedzający lub rok następujący po właściwym roku badania (1)-2-(1)	1. W każdym roku JPS dzielone na 4 podpróby • każdego roku jedna podpróba jest zastępowana przez nową	1. JPS dzielone są dla każdego kwartału na 4 próby elementarne, a następnie każda próba elementarna dzielona jest na 13 tygodniowych próbek elementarnych	1. Brak dzielenia próby
losowanie 783 JPS (dane dla 2014 r.)			losowanie 1350 JPS



# Losowanie JDS (1)

BBGD	EU.SILC	BAEL	ICT
W każdym JPS losuje się 24 mieszkania. Po 2 mieszkania na miesiąc.	<p>Liczba losowanych mieszkań w JPS zależna od rodzaju warstwy:</p> <ol style="list-style-type: none"> <li>1. miasta powyżej 100 tys. mieszkańców: 3 mieszkania</li> <li>2. miasta 20-100 tys.: 4 mieszkania</li> <li>3. miasta poniżej 20 tys.: 5 mieszkań</li> <li>4. tereny wiejskie: 6 mieszkań</li> </ol>	<p>Liczba mieszkań losowanych w każdym województwie proporcjonalna do pierwiasta z liczby mieszkańców w województwie (nadreprezentacja mniejszych województw)</p> <ul style="list-style-type: none"> <li>• tygodniowa próbka składa się z 44-90 mieszkań na województwo</li> <li>• 4-11 mieszkań w JPS w zależności od klasy miejscowości warstwy: <ul style="list-style-type: none"> <li>• tereny wiejskie i małe miasta: 8-9 mieszkań</li> <li>• średnie miasta: 6-7 mieszkań</li> <li>• tereny wielkomiejskie: 6 mieszkań</li> </ul> </li> </ul>	6 mieszkań w JPS

# Losowanie JDS (2)

BBGD	EU.SILC	BAEL	ICT
1. Losowanie proste - jednakowe prawdopodobieństwo wylosowania każdego mieszkania. Ma to przełożenie na schemat losowania JPS	1. Losowanie proste	1. Losowanie proste (?)	1. Losowanie proste (?)
1. model całkowitej rotacji z miesięcznym okresem wymiany próby. Każde mieszkanie badane jest w dwóch kolejnych latach	1. Mieszkanie jest badane w 4 kolejnych latach (co roku eliminowana jedna podpróba z czterech)	1. 4-krotne badanie jednego mieszkania (badanie w 2 następujących po sobie kwartałach - dwa kwartały przerwy - badanie w 2 następujących po sobie kwartałach)	1. bd
1. 37148 mieszkań (2015 r.)	1. ok 24000 mieszkań	1. 31,3 tys. gospodarstw domowych (w tym 20,5 tys. w miastach) oraz 65 tys. osób (w tym 38,8 tys. w miastach) - dane na 2 kwartału 2016	1. 8100 mieszkań

# Scematy losowania - informacje dodatkowe

## (1)

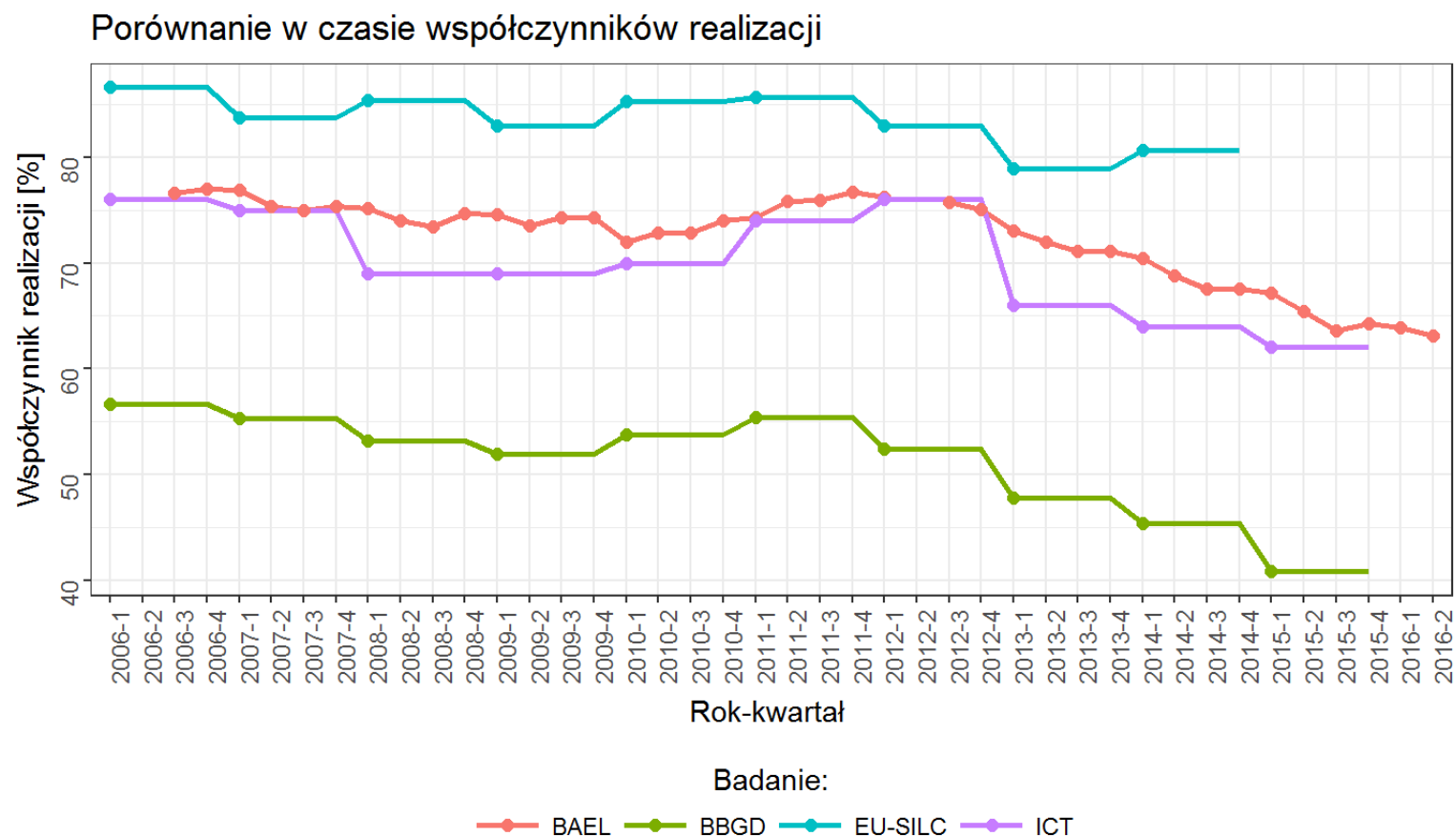
Wyszczególnienie	BBGD	EU.SILC	BAEL	ICT
Termin badania	Badanie ciągłe	5.05-18.07 (dla roku 2014)	Badanie ciągłe	Badanie coroczne
Forma badania	1. dwie dwutygodniowe książeczki (nt rozchdoow, przychodow, spożycia itp..) 2. wywiad bezpośredni na koniec miesiąca 3. wywiad kwartalny	wywiad bezpośredni		1. wywiad bezpośredni* * możliwe CATI
Czas badania	miesiąc + czas badania bezpośredniego na koniec miesiąca + czas badania kwartalnego	średnio 85 min. (dane z raportu 2014 r.)	bd	bd

# Schematy losowania - informacje dodatkowe

## (2)

	Wyszczególnienie	BBGD	EU.SILC	BAEL	ICT
4	Kwestionariusze	1. Kwestonariusz dla gospdoarstw domowych 2. Brak kwestionariusza dla osób indywidualnych	1. Kwestionariusz gospodarstw domowych • Kwestionariusz osób indywidualnych	1. Kwestionariusz dla osób indywidualnych	1. Kwestionariusz gospodarstw domowych • Kwestionariusz osób indywidualnych
5	Wiek osób badanych	(Do sprawdzenia)	16 lat i więcej	15 lat i więcej	1. Przed 2014 r.: 16-74 lata 1. Od 2014 r.: 12-74 lata

# Wskaźniki realizacji



**Wskaźnik realizacji** - udział gospodarstw domowych, które zostały przebadane i ukończyły badanie spośród wszystkich wylosowanych gospodarstw domowych kwalifikujących się do uczestniczenia w

# Braki odpowiedzi w przekrojach

Nazwa	BBGD 2006-12	BBGD 2013-15	EU-SILC	BAEL	ICT
wskaźnik realizacji		TAK	TAK	TAK	TAK
wskaźnik braków odpowiedzi	TAK		TAK		
wskaźnik częściowego braku odpowiedzi	TAK	TAK	TAK		
wskaźnik całkowitego braku odpowiedzi	TAK		TAK		
próba podstawowowa/rezerwowa		TAK			
województwa		TAK		TAK	
klasy miejscowości		TAK		TAK	
kwartały	TAK			TAK	
który raz w badaniu	TAK	TAK			
grupy społeczno-ekonomiczne	TAK				
przyczyny nieuczestniczenia	TAK	TAK	TAK		
wielkość gospodarstwa domowego	TAK				
wskaźniki dla osób indywidualnych			TAK		TAK
Regiony					
Powiaty					

# Przekroje braków odpowiedzi, a przekroje szacunków

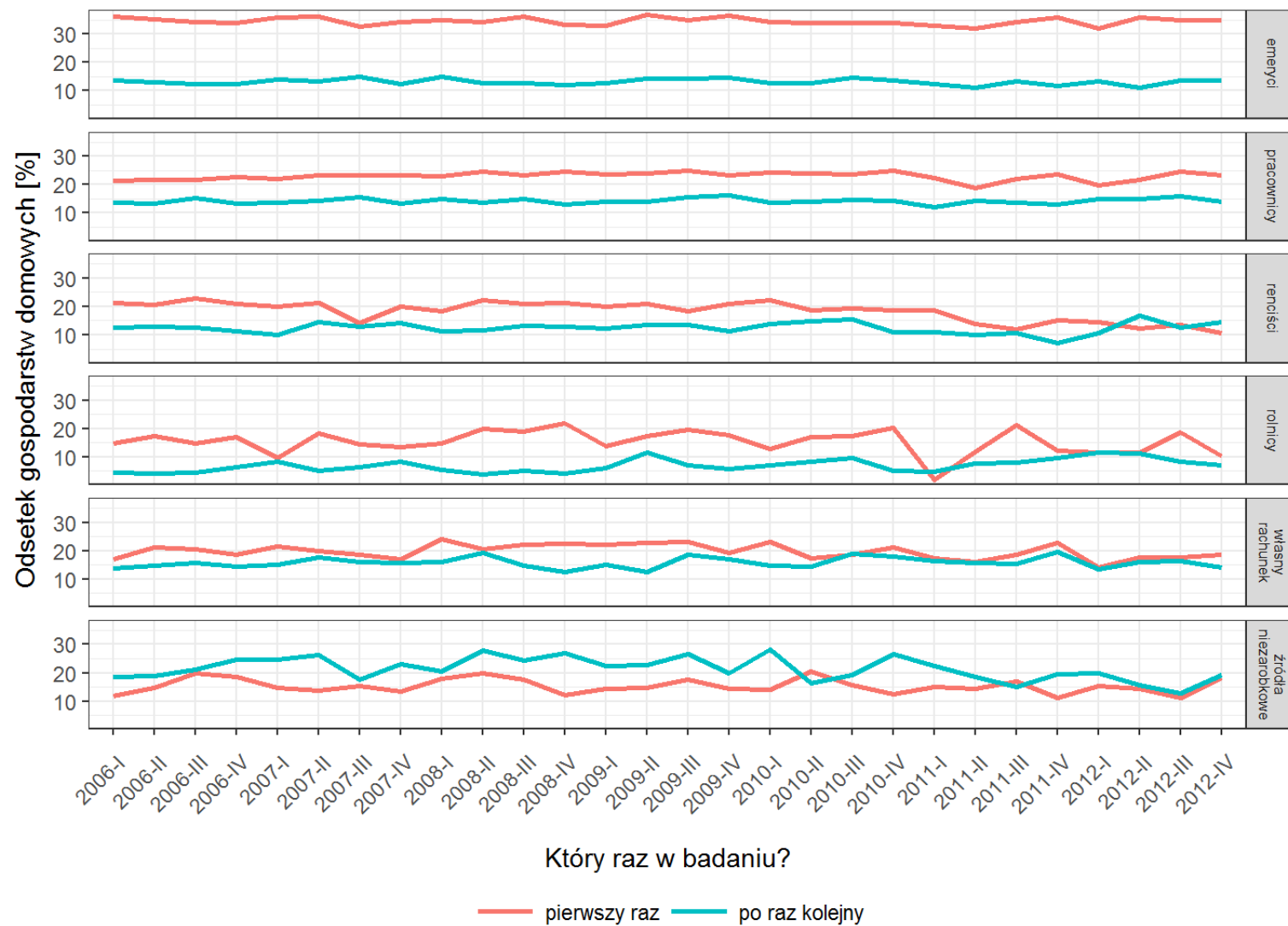
Badanie Budżetów Gospodarstw Domowych 2013-2015

Nazwa	Braki odpowiedzi	Szacunki
próba podstawowa/rezerwowa	TAK	
województwa	TAK	TAK
klasy miejscowości	TAK	TAK
który raz w badaniu	TAK	
grupy społeczno-ekonomiczne		TAK
przyczyny nieuczestniczenia	TAK	
kwantyle dochodu		TAK
typ biologiczny		TAK
wykształcenie		TAK
rodzaje budynków		TAK

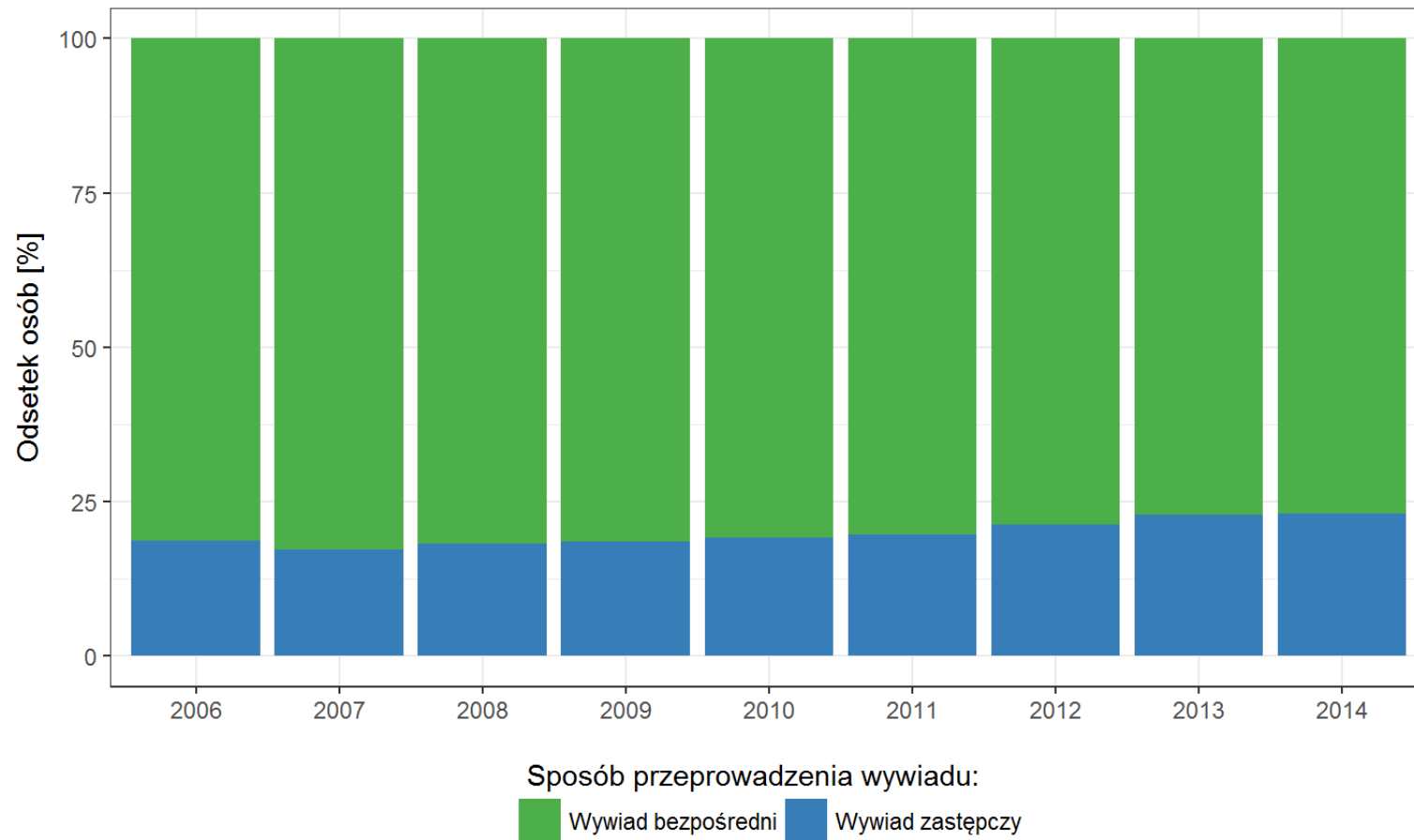
# Przykłady



# Braki odpowiedzi w BBGD w zależności od momentu dołączenia do badania w latach 2006-2012



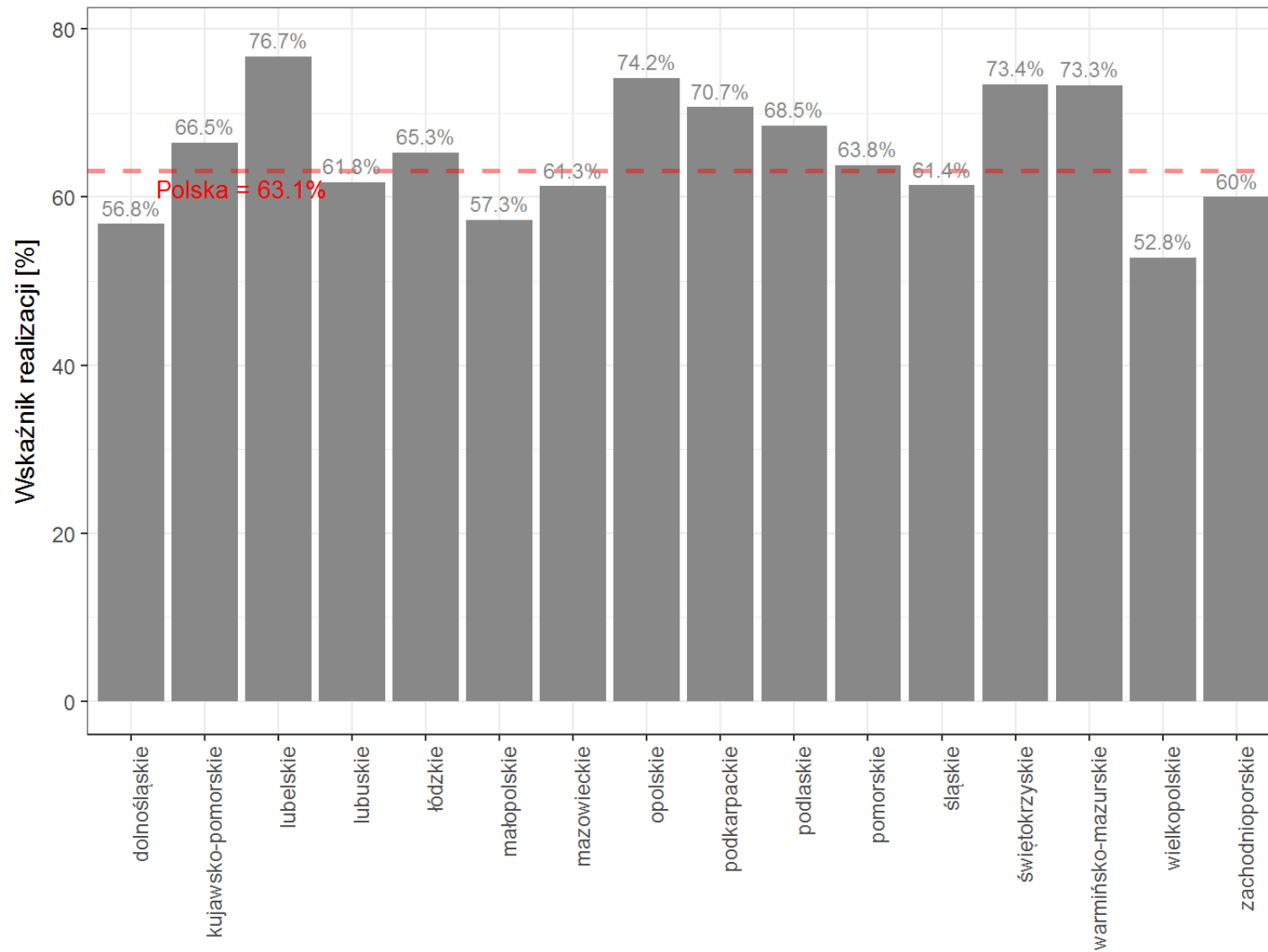
Przeprowadzone wywiady według rodzaju w badaniu EU-SILC  
w latach 2006-2014



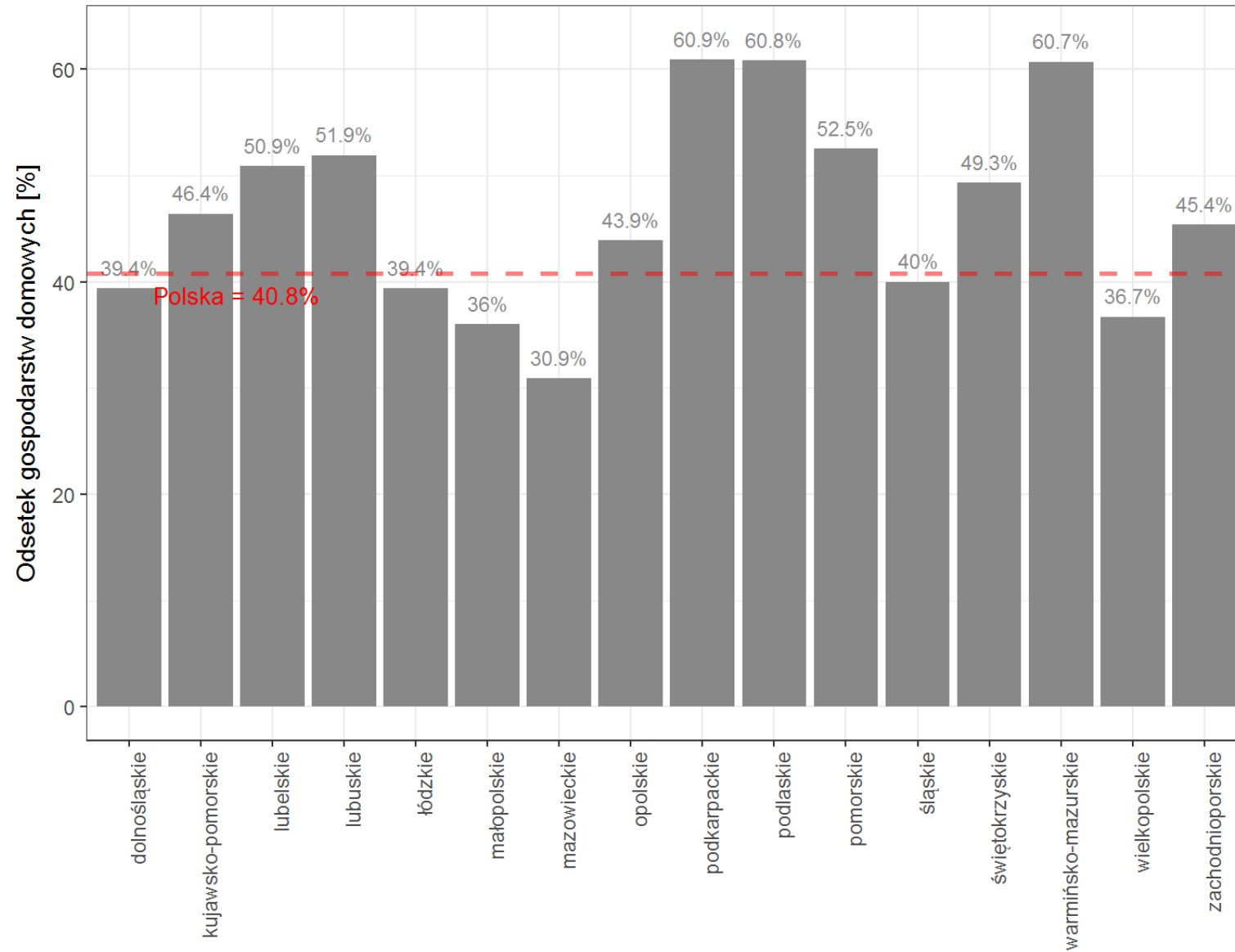
**Wywiad bezpośredni** - wywiad udzielany przez osobę, której dotyczy badanie

**Wywiad zastępczy** - wywiad udzielany przez osobę 'trzecią', która jest w stanie udzielić rzetelnych

Wskaźnik realizacji dla BAEL według województw  
w 2. kwartale 2016 r.



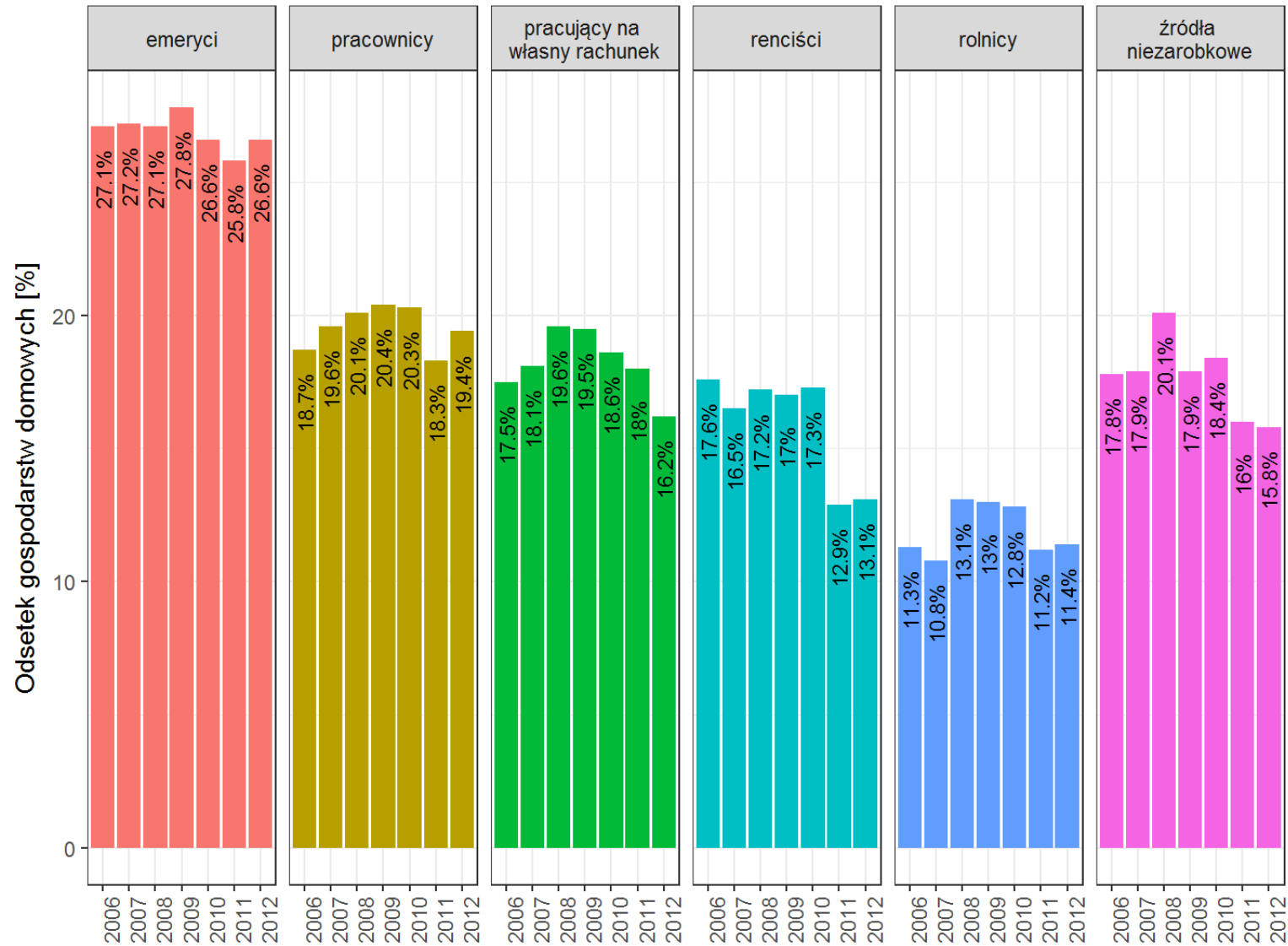
Wskaźnik realizacji BBGD według województw w 2016 r.



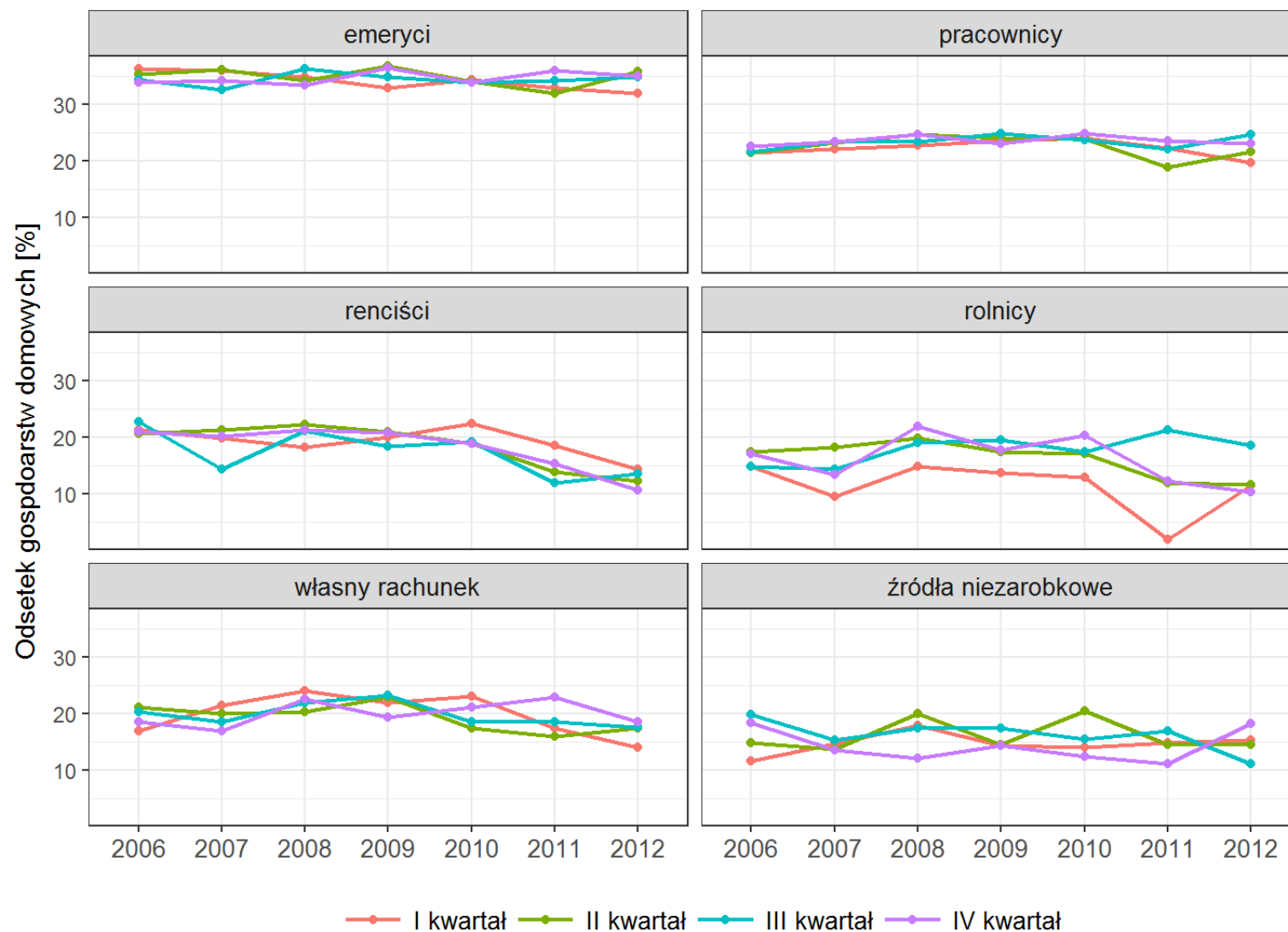
# Kwartyle wskaźników realizacji

# Wskaźniki realizacji według klas miejscowości

Braki odpowiedzi według grup społeczno-ekonomicznych  
w BBGD w latach 2006-2012

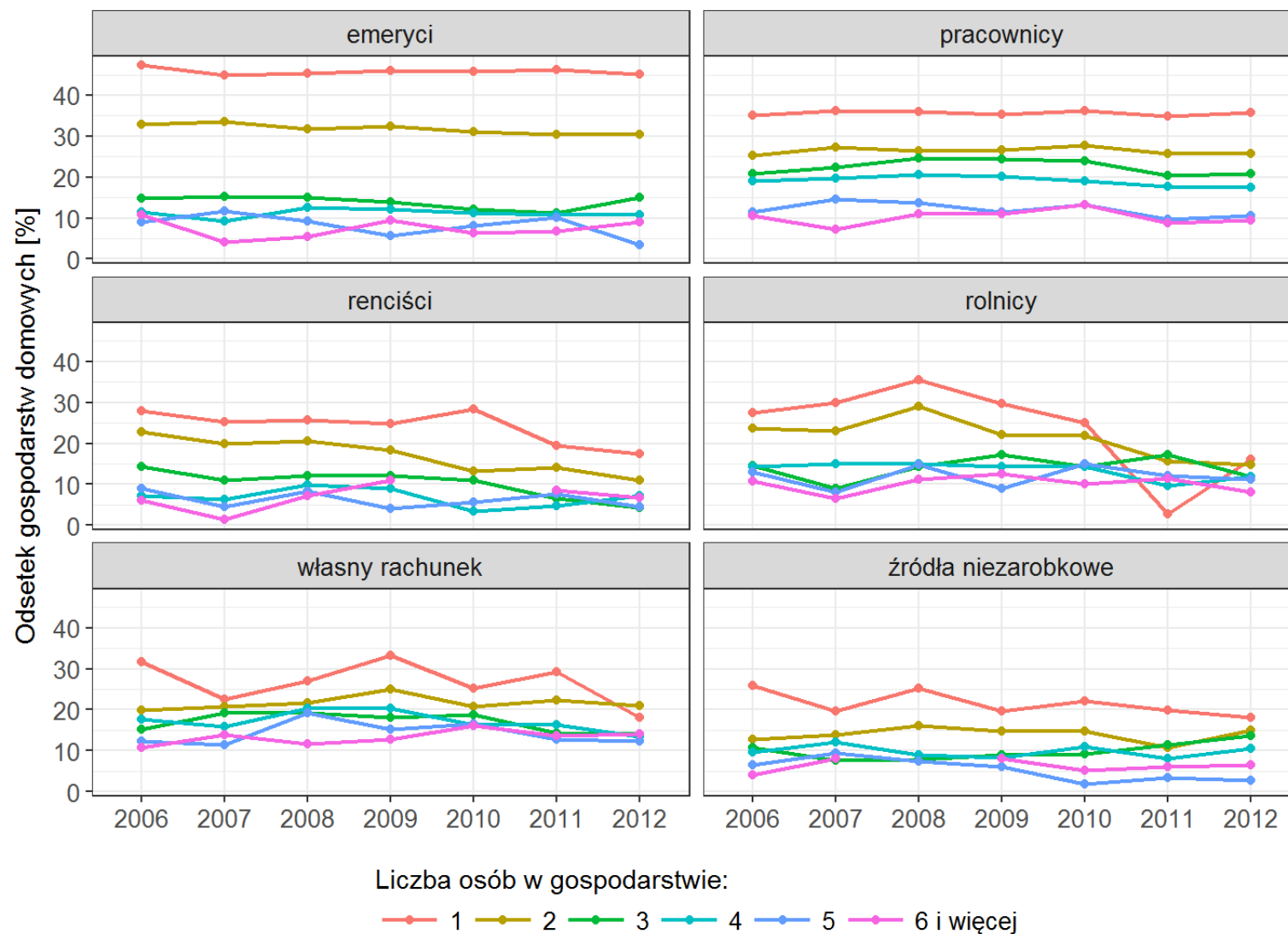


Braki odpowiedzi w grupach społeczno-ekonomicznych według kwartałów w BBGD  
(GD biorące udział w badaniu po raz pierwszy)





Odmowy odpowiedzi w grupach społeczno-ekonomicznych według wielkości gospodarstwa domowego w BBGD (GD biorące udział w badaniu po raz pierwszy)



# Podsumowanie

1. Braki odpowiedzi są istotnym czynnikiem mogącym wpłynąć na wyniki analiz, raportów itp.
2. Informacje na temat braków odpowiedzi w badaniu powinny być jednymi z kluczowych danych zawartych w raporcie
3. Każde z analizowanych badań posiada informacje na temat braków danych
4. Badania charakteryzują się zróżnicowaną tematyką i sposobem organizacji, co determinuje zakres podawanych informacji, w tym tych o brakach odpowiedzi
5. W poszczególnych badaniach informacje te są zróżnicowane pod względem formy, obszerności oraz szczegółowości podawanych treści
6. Istnieje dość silna korelacja między wskaźnikami realizacji analizowanych badań
7. Na podstawie analizy zgromadzonych danych i materiałów można określić pewne wzory (prawidłowości) w występowaniu braków odpowiedzi w różnych przekrojach

# Rekomendacje

1. Zdefiniowanie pojęcia 'brak odpowiedzi'
2. Wypracowanie jednolitego nazewnictwa (wskaźnik realizacji, współczynnik realizacji, wskaźnik kompletności)
3. Ujednolicenie przekrojów dla jakich podawane są informacje na temat braków odpowiedzi - umożliwienie porównania między badaniami
4. Rozszerzenie przekrojów o np. regiony, podregiony, powiaty
5. Dostosowanie przekrojów dla braków danych i szacunków (na ile to możliwe)
6. Podawanie danych na temat braków odpowiedzi dla poziomu całego kraju w przekrojach 'nietrytorialnych' (taka sytuacja nie zawsze ma miejsce)
7. Zachowanie ciągłości podawanych informacji o brakach odpowiedzi w ramach badania - dodawanie nowych obszarów informacyjnych zamiast ich zastępowania (przykład BBGD)
- 8.

# Literatura

1. Publikacje raportów z badania Budżety Gospodarstw Domowych za lata 2006-2015
2. Publikacje raportów z badania Dochody i Warunki życia Ludności Polski (raport z badania EU-SILC za lata 2006-2014)
3. Publikacje raportów z badania Społeczeństwo Informacyjne w Polsce. Wyniki badań statystycznych z lat 2006-2015
4. Publikacje badania Aktywność Ekonomiczna Ludności Polski za okres od 1 kwartału 2001 r. do 2 kwartału 2016 r.
5. Szymkowiak, M., 2009, Estymatory kalibracyjne w badaniu Budżetów Gospodarstw Domowych, Poznań, s. 9-12
6. Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys, 2016, The American Association for Public Opinion Research, s. 23-27, [http://www.aapor.org/AAPOR\\_Main/media/publications/Standard-Definitions20169theditionfinal.pdf](http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf)