

FlatironKitchen

How we overhauled a Frankensteinian workflow with the tidyverse to enable fast, reproducible, elegant analyses of electronic health records.

rstudio::conf 2020, San Francisco

Nathaniel D. Phillips

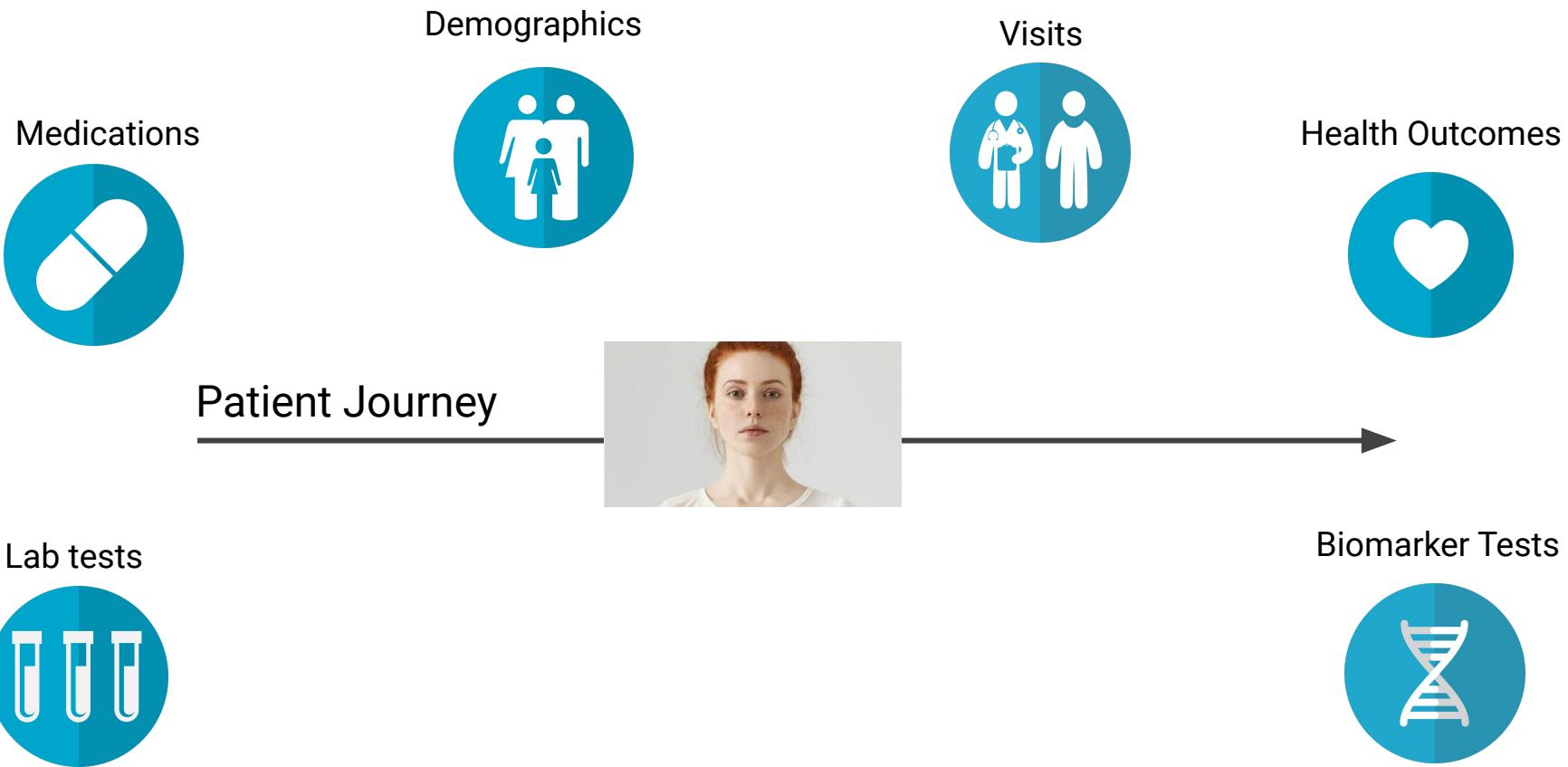
On behalf of Roche / Genentech PHC Data Science Analytics

✉ phc.datascience@roche.com

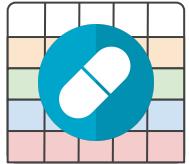


Patient Journey

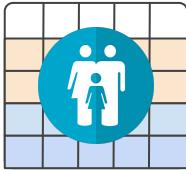




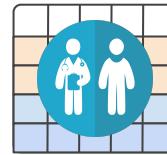
Medications



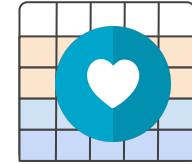
Demographics



Visits



Health Outcomes



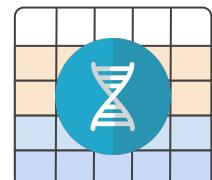
Patient Journey

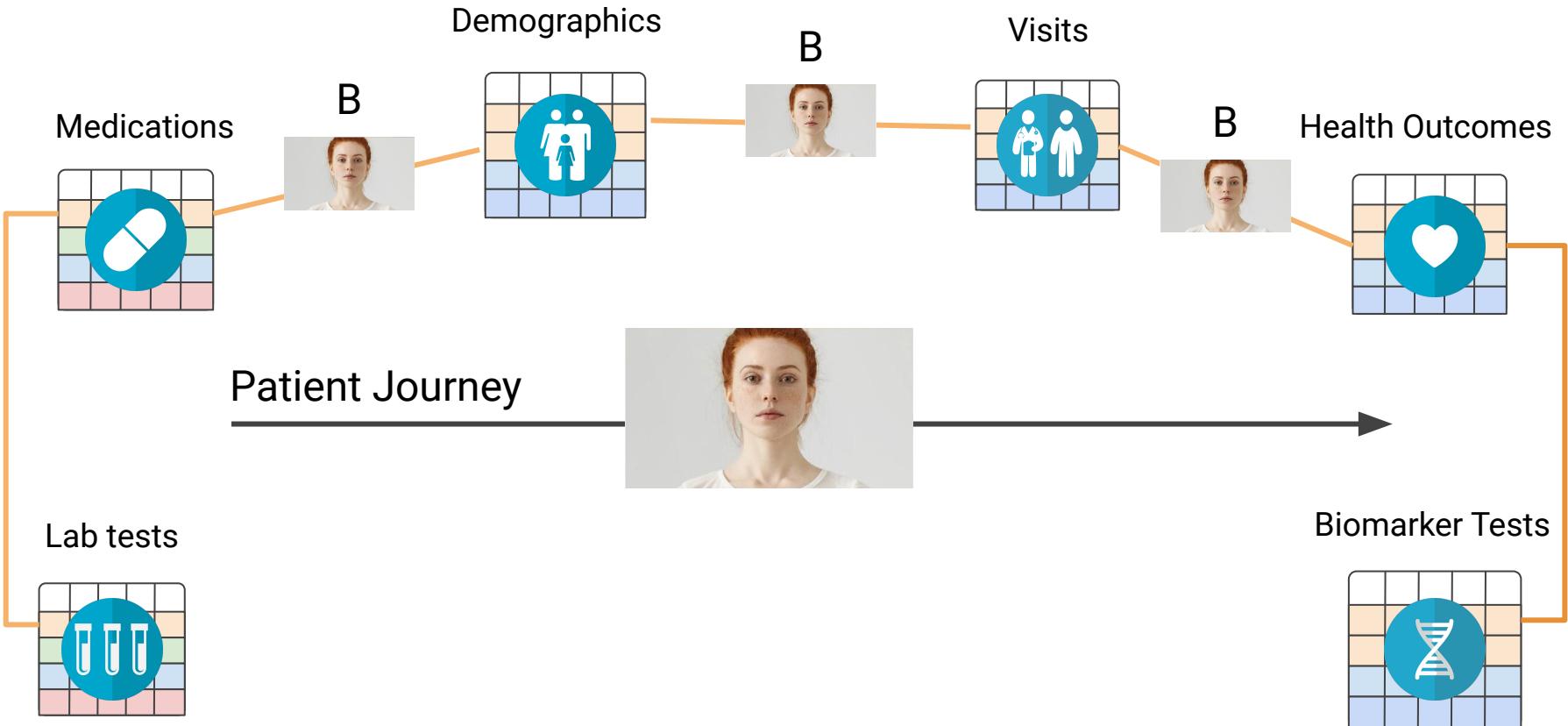


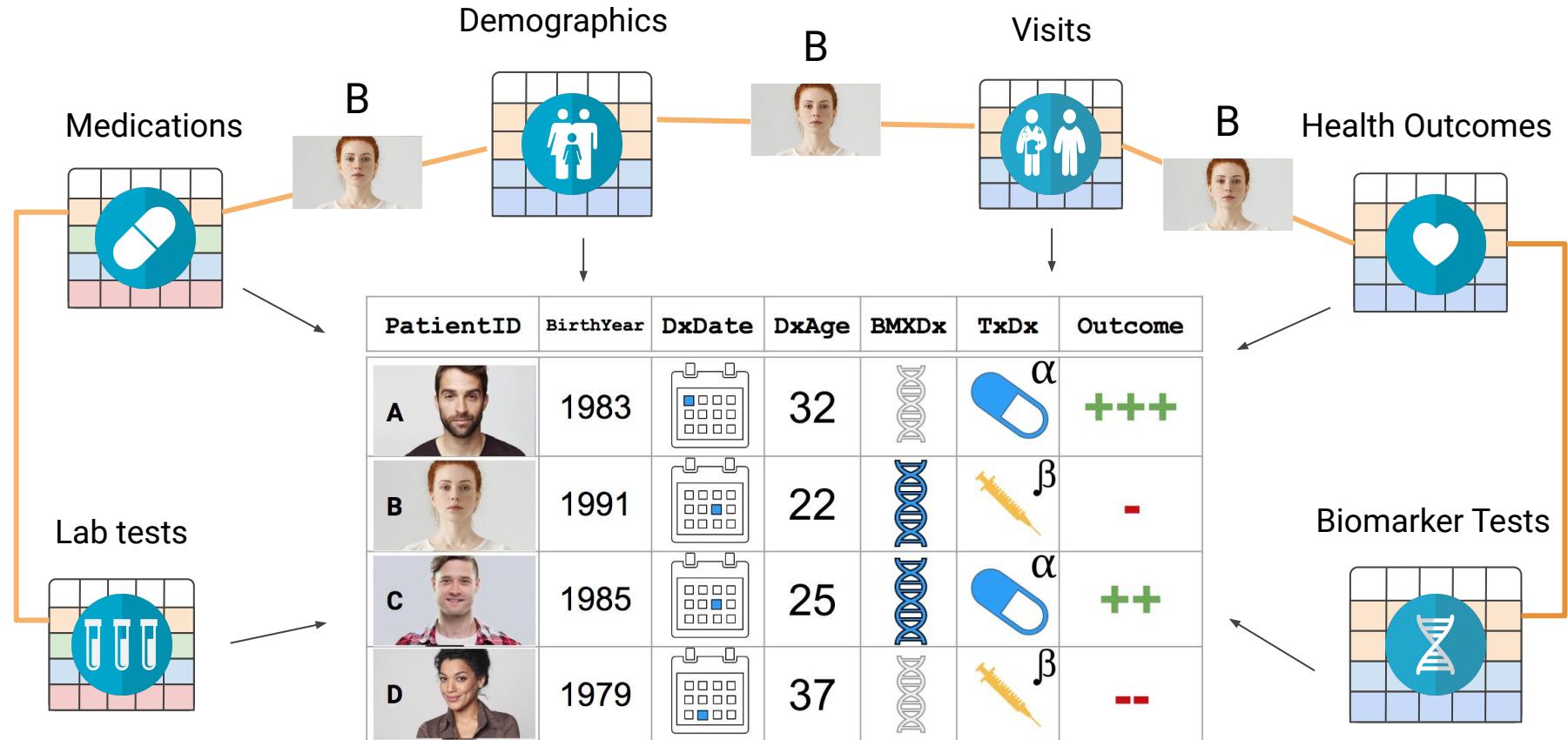
Lab tests



Biomarker Tests









Tidy, Analysis-Specific One-Row-Per-Patient (ORPP) Dataframe

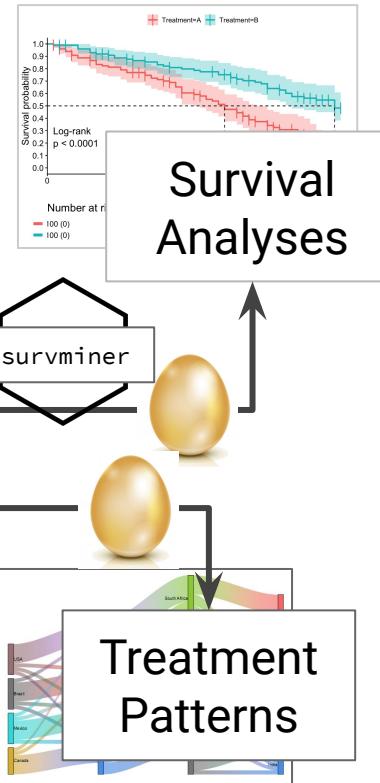
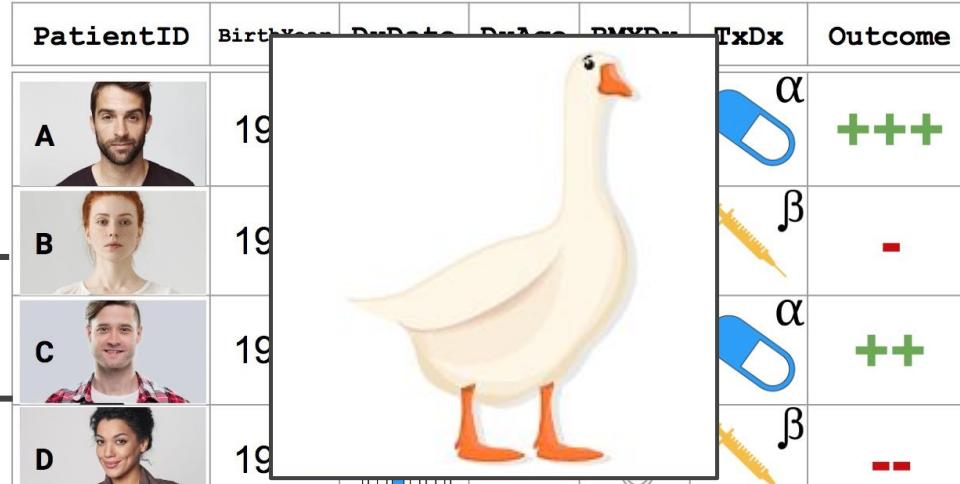
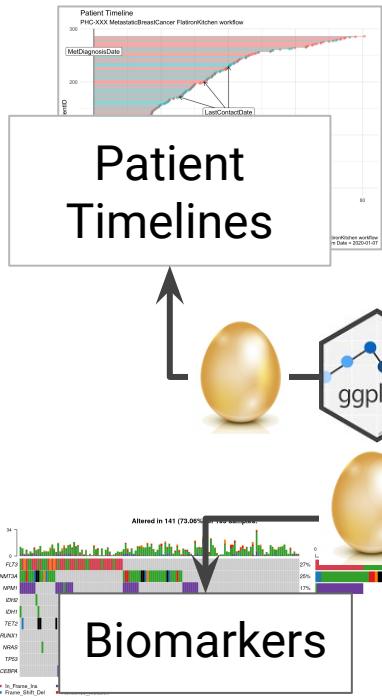
PatientID	BirthYear	DxDate	DxAge	BMxDx	TxDx	Outcome
A	1983		32		α	+++
B	1991		22		β	-
C	1985		25		α	++
D	1979		37		β	--



Tidy, Analysis-Specific One-Row-Per-Patient (ORPP) Dataframe

PatientID	BirthYear	DiagDate	DiagYear	TxDx	Outcome
A	19			α	+++
B	19			β	-
C	19			α	++
D	19			β	--

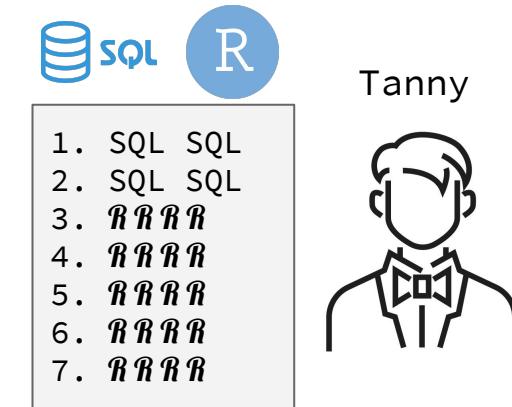
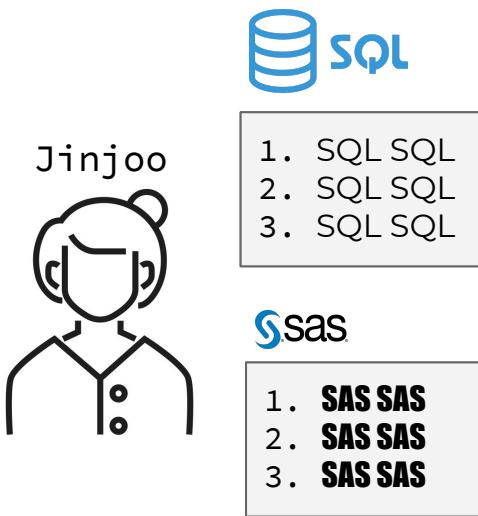
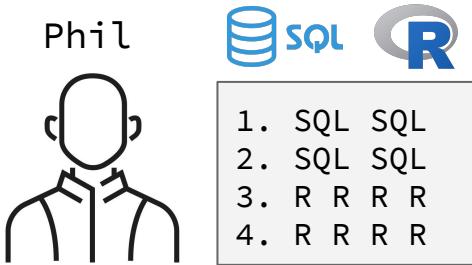
Tidy, Analysis-Specific **One-Row-Per-Patient (ORPP)** Dataframe



Note: All code and outputs are for illustration purposes only, may have been edited for clarity and confidentiality, and do not accurately reflect real patient data.

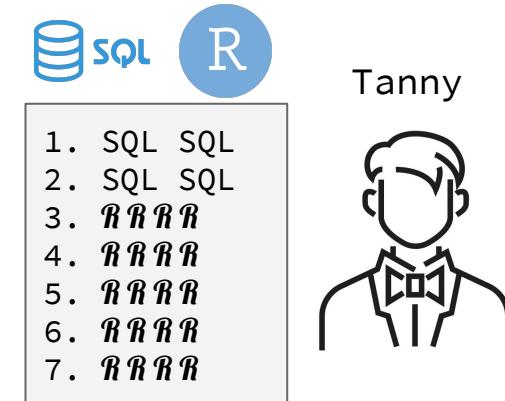
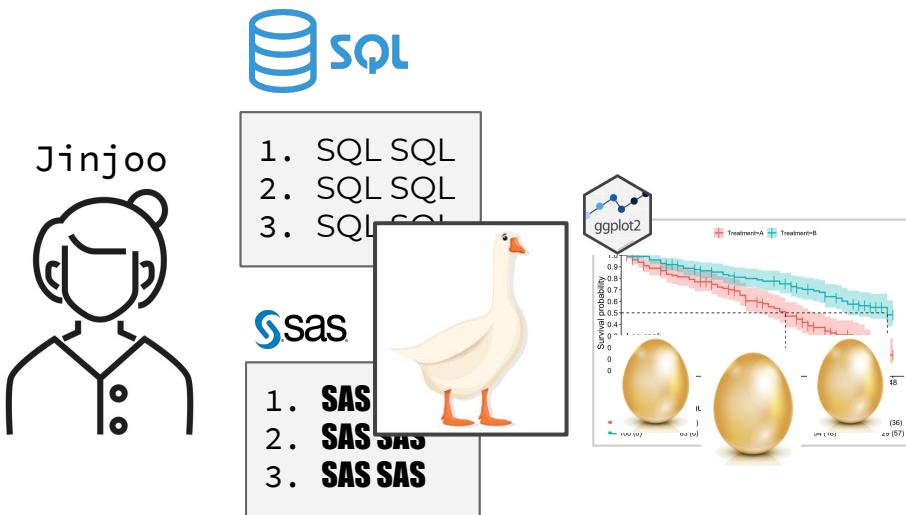
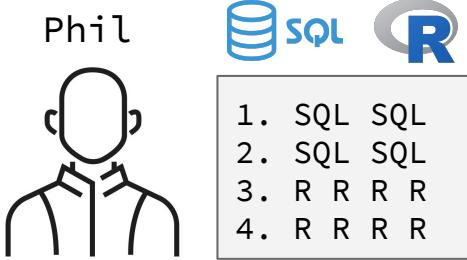
How we **were** doing it

Static, siloed, inconsistent



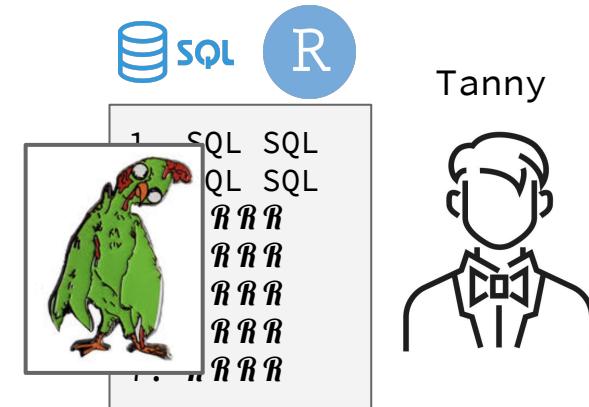
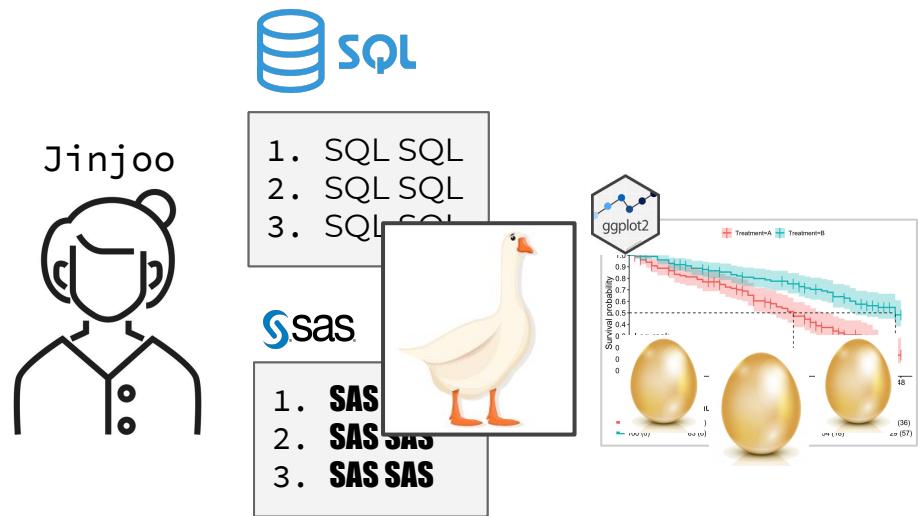
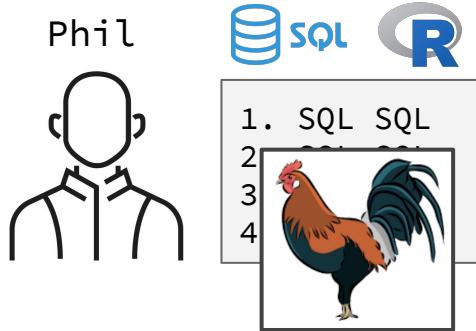
How we **were** doing it

Static, siloed, inconsistent



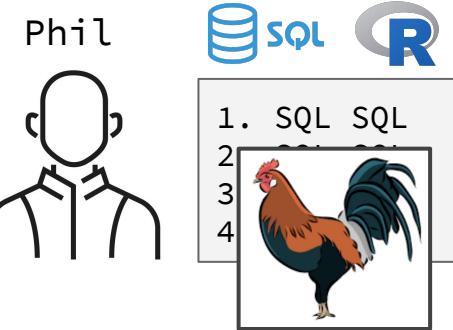
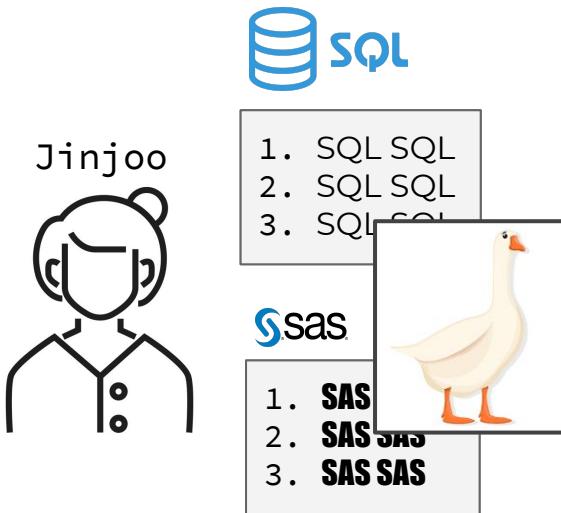
How we **were** doing it

Static, siloed, inconsistent



How we **were** doing it

Static, siloed, inconsistent

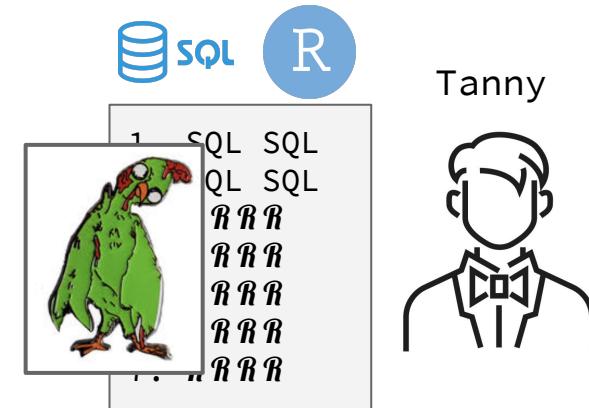


PROBLEMS

- ✗ High Code Duplication
- ✗ Poor / No Testing
- ✗ Poor / No Documentation
- ✗ Inconsistent Style

RESULT

- ✗ High QC Costs
- ✗ Intimidate and burden newcomers

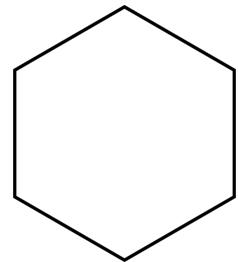


How we **wanted** to do it

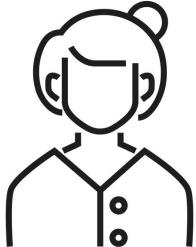
Functional,
transparent, tested.



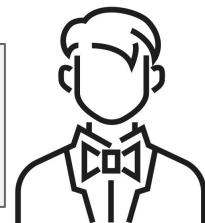
Phil



Jinjoo

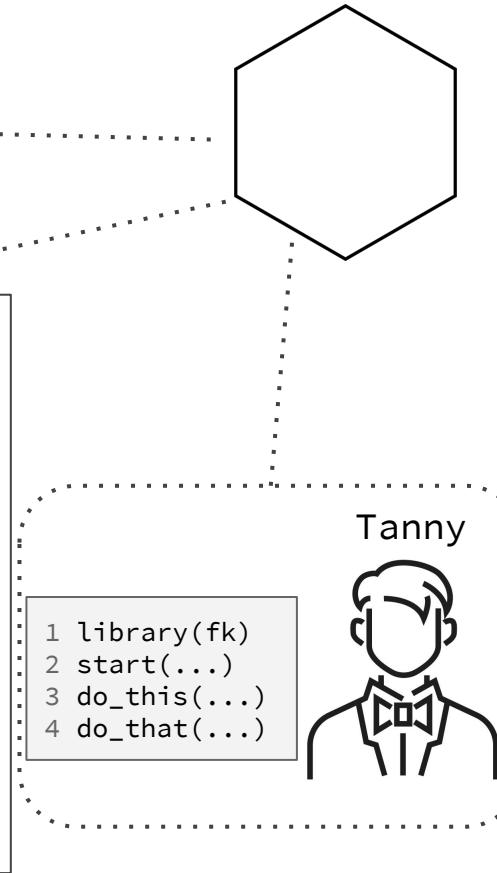
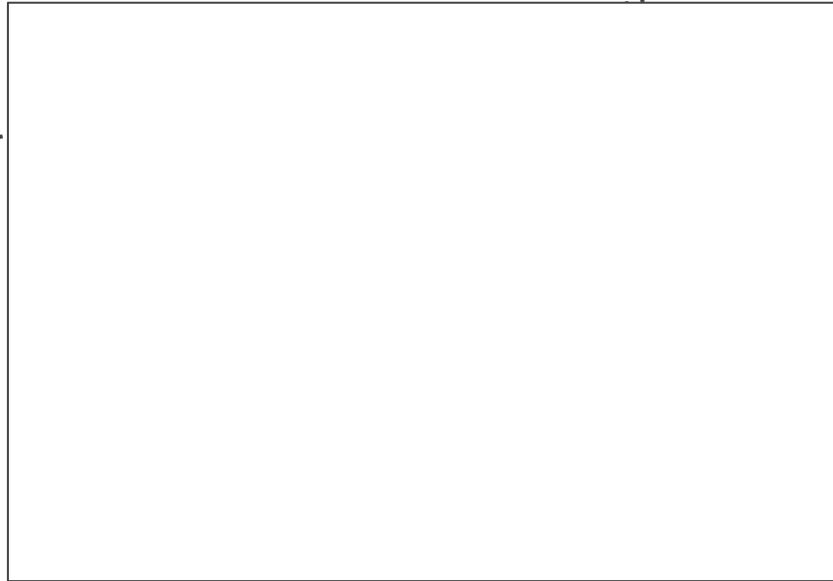
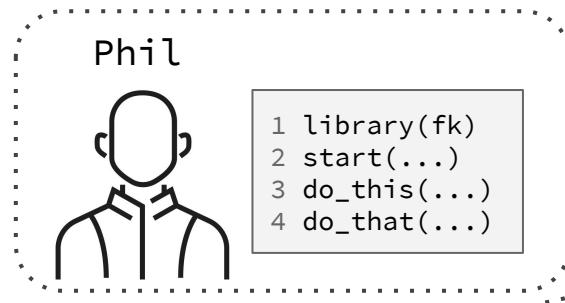


Tanny



How we **wanted** to do it

Functional,
transparent, tested.



How we **wanted** to do it

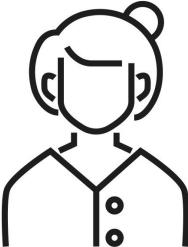
Functional,
transparent, tested.

Phil



```
1 library(fk)
2 start(...)
3 do_this(...)
4 do_that(...)
```

Jinjoo



```
1 library(fk)
2 start(...)
3 do_this(...)
4 do_that(...)
```

Modular, Readable

```
start_this(...)
do_this(...)
do_that(...)
plot_this(...)
```

End-To-End “Recipes”

Advanced NSCLC Example

Nathaniel Phillips (nathaniel.phillips@roche.com)

2020-01-08



Documented

```
?do_this(...)
```

*This function
does...*

Tested and verbose

Ok!

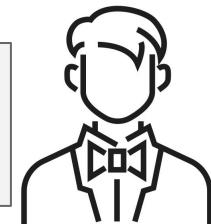
Note

Warning

STOP!

: ...
: ...
: ...
: ...

Tanny

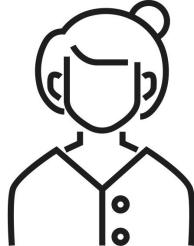


```
1 library(fk)
2 start(...)
3 do_this(...)
4 do_that(...)
```

How we **wanted** to do it

Functional,
transparent, tested.

Jinjoo



```
1 library(fk)
2 start(...)
3 do_this(...)
4 do_that(...)
```

Phil



```
1 library(fk)
2 start(...)
3 do_this(...)
4 do_that(...)
```

Modular, Readable

```
start_this()
do_this(...)
do_that(...)
plot_this()
```

Documented

?do_this(...)

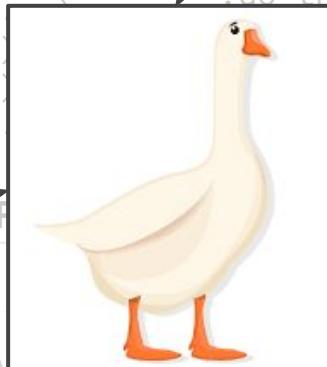
unction

End-To-End R

Advanced NSCLC Example

Nathaniel Phillips (nathanphillips@roche.com)

2020-01-08



STOP!

and verbose

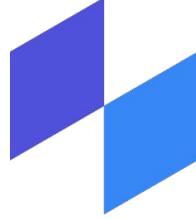
```
1 library(fk)
2 start(...)
3 do_this(...)
4 do_that(...)
```

Tanny



Invite and **enable** new-comers

Note: All code and outputs are for illustration purposes only, may have been edited for clarity and confidentiality, and do not accurately reflect real patient data.



flatiron



FlatironKitchen

Originally designed for **real-world oncology** data from **Flatiron Health**.

Tools, recipes, and a communal place to create and enhance patient cohorts.



Disclaimer

All subsequent code and outputs are **illustrative** of FlatironKitchen functionality. They are not verbatim.

Results do not accurately reflect real patient data from Flatiron Health or any other source.

FlatironKitchen is (as of January 2020)
not publically available

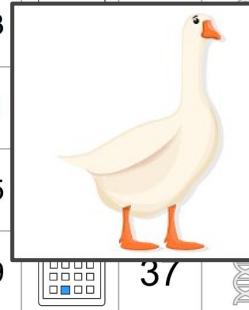
FlatironKitchen Objects



fk\$data

Analysis-ready,
One-Row-Per-Patient dataframe

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983				α	+++
B	1991				β	-
C	1985				α	++
D	1979		37		β	--



```
library(FlatironKitchen)

fk <- fi_start(...)

class(fk)
# [1] "FlatironKitchen"

names(fk)
# [1] "data" "connection"
# [3] "patients" "results"
#       [...]
```



fk\$connection

Connection to a database
(or local folder)



fk\$results

Cohort attrition, logs, warnings,
statistical objects



fk\$patients

Patient level meta-data (e.g.;
unique IDs, N, ...)



fk\$stored

Local copies of raw tables for
easy access

Note: All code and outputs are for illustration purposes only, may have been edited for clarity and confidentiality, and do not accurately reflect real patient data.



FlatironKitchen Objects

FlatironKitchen objects are of S3 class.

```
library(FlatironKitchen)

fk <- fi_start(...)

class(fk)
# [1] "FlatironKitchen"

names(fk)
# [1] "data" "connection"
# [3] "patients" "results"
#       [...]
```

```
print(fk)
```

Nathaniel's 28-01-2020 Skin Cancer Analysis

Project

```
$version : 0.2.3 'Roeschti'  
$datamart: "BreastCancer"  
$delivery: 2019-12 "December 2019"
```

Patients

```
$n      : 10,000  
$ids    : [A, B, D, E, ...]
```

Attrition

```
All BreastCancer Patients   : 100,000 -----  
Diagnosed after 20XX        : 80,000 -----  
Given drug X in LOT1        : 30,000 -----  
Biomarker X positive at DX : 10,000 --
```

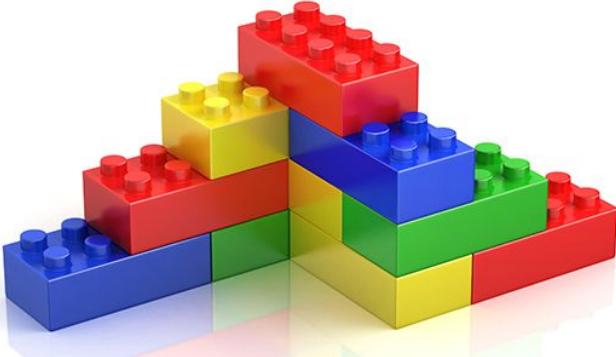
Warnings

- 2 patients had a recorded event after death.
- 5 patients have no recorded visits after diagnosis.
- ...

FlatironKitchen Functions



Functions to create, enhance,
and analyse a cohort



fk\$data

Analysis-ready,
One-Row-Per-Patient dataframe

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983		32		α	
B	1991		22		β	-
C	1985		25		α	
D	1979		37		β	--

FlatironKitchen Functions



fi_cohort_x()

fi_cohort_start(),
fi_cohort_include(), ...

Start a cohort or remove patients from an existing one.

"Start a patient cohort containing patients with a Skin Cancer diagnosis"

fi_cohort_start("SkinCancer")

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983		32			
B	1991		22			
C	1985		25			
D	1979		37			

--- fi_cohort_start() -----

Touching: DIAGNOSIS

NewCols : PatientID [c], Diagnosis [c],
DiagnosisDate [d]

Cohort : 1,000

FlatironKitchen Functions



fi_add_x()

fi_add_demographics()
fi_add_biomarkers(), ...

Add (join) variables from external tables

"Add demographics information"

fi_add_demographics()

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983		32		α	+++
B	1991		22			
C	1985		25			
D	1979		37			

--- fi_add_demographics() ---

Touching: DEMOGRAPHICS

NewCols : Sex [c], BirthYear [n], Race [c],
...

FlatironKitchen Functions



fi_add_x()

fi_add_demographics()
fi_add_biomarkers(), ...

Add (join) variables from external tables

"Add mortality information"

fi_add_mortality()

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983	📅	32	DNA	α	+++
B	1991	📅	22	DNA		
C	1985	📅	25	DNA		
D	1979	📅	37	DNA		

A light blue cloud-like shape surrounds the table, containing icons representing various types of data: a calendar, DNA double helix, a blue capsule, and a large blue heart.

--- fi_add_mortality() ---

Touching: MORTALITY

Note : Death = !is.na(DateOfDeath)

NewCols : DateOfDeath [d], Death [l]

Note : Death info found for 100 / 1,000
(10%) of patients

FlatironKitchen Functions



fi_add_x()

```
fi_add_demographics()  
fi_add_biomarkers(), ...
```

Add (join) variables from external tables

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983		32		α	+++
B	1991		22			
C	1985		25			
D	1979		37			



"Add **last value** of Biomarker **X** observed up to **1 year prior to diagnosis date**"

```
fi_add_biomarker(  
  which = "X", index = "DiagnosisDate",  
  left = -365, right = 0, method = "last")
```

--- fi_add_biomarker() -----

Explain : Adding **last known value** of biomarker X observed in [-365, 0] days of DiagnosisDate

Touching: **BIMARKERS**

NewCols : BiomarkerXDiagnosisLast [c]

Note : Biomarker X values found for 500 / 1,000 (50%) of patients

FlatironKitchen Functions



fi_add_x()

fi_add_demographics()
fi_add_biomarkers(), ...

Add (join) variables from external tables

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983		32		α	+++
B	1991		22			
C	1985		25			
D	1979		37			



"Add **number** of tests of Biomarker **X** observed up to 1 year prior to diagnosis date"

```
fi_add_biomarker(  
  which = "X", index = "DiagnosisDate",  
  left = -365, right = 0, method = "N")
```

--- fi_add_biomarker() -----

Explain : Adding **number of tests** of biomarker X observed in [-365, 0] days of DiagnosisDate

Touching: **BIMARKERS**

NewCols : BiomarkerXDiagnosisN [n]

Note : >= 1 Biomarker X tests found for 500 / 1,000 (50%) of patients

FlatironKitchen Functions



fi_calc_x()

fi_calc_age(),
fi_calc_duration(), ...

Calculate new variables from existing
.data object

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983		32			α +++
B	1991		22			β -
C	1985		25			α ++
D	1979		37			β --

"Calculate **age** at **diagnosis** and **death**

```
fi_calc_age()  
when = c("DiagnosisDate", "DateOfDeath"),  
...)
```

--- fi_calc_age() -----

Note : Calculating patient age at DiagnosisDate, DateOfDeath and imputing BirthDate as 02-July-BirthYear

NewCols : DiagnosisAge [n], DeathAge [n]

Warning : 15 / 1,000 (1.5%) of patients had a negative calculated age!

FlatironKitchen Functions



fi_cohort_x()

fi_cohort_start(),
fi_cohort_include(), ...

Start a cohort or remove patients from an existing one.

PatientID	BirthYear	DxDate	DxAge	BMxDx	TxDx	Outcome
A	1983		32			α
B	1991		22			β
C	1985		25			α
D	1979		37			β

"Apply inclusion criteria: Patients diagnosed after 2012"

```
fi_cohort_include(  
  DiagnosisDate > "2012-12-31",  
  why = "Must be diagnosed after 2012")
```

--- fi_cohort_include() -----

Note : Applying Inclusion Criteria =
 : 'Must be diagnosed after 2012'
 : Code = DiagnosisDate > "2012-12-31"
Excluded: 100 patients
Cohort : 900 patients

FlatironKitchen Functions



fi_cohort_x()

fi_cohort_start(),
fi_cohort_include(), ...

Start a cohort or remove patients from an existing one.

PatientID	BirthYear	DxDate	DxAge	BMxDx	TxDx	Outcome
A	1983		32			α
B	1991		22			β
C	1985		25			α
D	1979		37			β

"Apply inclusion criteria: Follow up time (from Diagnosis to Last Contact) is positive"

```
fi_cohort_include(  
  FollowUpDays > 0,  
  why = "Positive followup time")
```

--- fi_cohort_include() -----

Note : Applying Inclusion Criteria =
 : 'Positive followup time'

Excluded: 50 patients

Cohort : 850 patients

FlatironKitchen Functions



Our final “Recipe”



```
1 fk <- fi_cohort_start("SkinCancer") %>%
2     fi_add_demographics(...) %>%
3     fi_add_mortality(...) %>%
4     fi_add_lastcontact(...) %>%
5     fi_add_biomarker(...) %>%
6     fi_add_biomarker(...) %>%
7     fi_calc_age(...) %>%
8     fi_calc_duration(...) %>%
9     fi_cohort_include(...) %>%
10    fi_cohort_include(...) %>%
11    fi_cohort_include(...)
```

FlatironKitchen Functions

Our final “Recipe”

```
1 fk <- fi_cohort_start("SkinCancer") %>%
2   fi_add_demographics(...) %>%
3   fi_add_mortality(...) %>%
4   fi_add_lastcontact(...) %>%
5   fi_add_biomarker(...) %>%
6   fi_add_biomarker(...) %>%
7   fi_calc_age(...) %>%
8   fi_calc_duration(...) %>%
9   fi_cohort_include(...) %>%
10  fi_cohort_include(...) %>%
11  fi_cohort_include(...)
```



vs



```
SELECT emprera.distribution_code, bbase.name AS DistributionName,
       COUNT(DISTINCT raises.person_id) AS NumRaises,
       COUNT(DISTINCT lastraise.person_id) AS NumLastRaises,
       SUM(bases.annual_rate) / (SUM(bases.annual_rate) +
       CONVERT(DECIMAL(10, 0), (SUM(raises.annual_rate) -
       CONVERT(DECIMAL(10, 0), MAX(raises.annual_rate) * baseerate.annual_rate)) * Mins) AS AvgRaiseRate,
       CONVERT(DECIMAL(10, 0), MIN(bases.annual_rate - baseerate.annual_rate)) AS MinRate
  FROM emprera.distribution_code
  INNER JOIN emprera.company_code ON emprera.company_code = raises.company_code
  INNER JOIN emprera.person ON raises.person_id = emprera.person_id
  LEFT OUTER JOIN lastraise ON raises.person_id = lastraise.person_id
  AND bbase.distribution_code = emprera.distribution_code
  INNER JOIN emprera.baseerate ON lastraise.person_id = baseerate.person_id
  AND baseerate.person_id = raises.person_id
  AND baseerate.effective_date = lastraise.effective_date
  AND baseerate.company_code = lastraise.company_code
  FROM emprera.lastraise
 WHERE lastraise.person_id = baseerate.person_id
 AND lastraise.distribution_code = baseerate.distribution_code
 AND lastraise.effective_date <= (SELECT StartDate FROM emprera.distribution_code)
 WHERE raises.company_code = (Input Company Code)
 AND raises.effective_date <= (Input End Date)
 AND raises.effective_date >= (SELECT MaxLastRaiseEffectiveDate)
 WHERE lastraise.person_id = raises.person_id
 AND lastraise.distribution_code = raises.distribution_code
 AND lastraise.effective_date <= (Input End Date)
 GROUP BY emprera.distribution_code, bbase.name
 ORDER BY emprera.distribution_code
```

```
SELECT XX XXX X
INNER JOIN XXXXX X
WHERE XX XXX X
AND, ON, IF
GROUP BY XX XXX
CREATE XX XXX X
ORDER BY XX XXX X
```



query_1.sql
query_2.sql
setup.R
wrangle.R
...

FlatironKitchen Functions

Our final “Recipe”

```
1 fk <- fi_cohort_start("SkinCancer") %>%
2   fi_add_demographics(...) %>%
3   fi_add_mortality(...) %>%
4   fi_add_lastcontact(...) %>%
5   fi_add_biomarker(...) %>%
6   fi_add_biomarker(...) %>%
7   fi_calc_age(...) %>%
8   fi_calc_duration(...) %>%
9   fi_cohort_include(...) %>%
10  fi_cohort_include(...) %>%
11  fi_cohort_include(...)
```



Skin Cancer Analysis

Jinjoo Shim (jinjoo.shim@roche.com)

Goal

Understand patient outcomes diagnosed with Skin Cancer...

Cohort Creation

```
library(FlatironKitchen)

fk <- fi_cohort_start("SkinCancer") %>%
  fi_add_demographics(...) %>%
  fi_add_mortality(...) %>%
  fi_add_lastcontact(...) %>%
  fi_add_biomarker(...) %>%
  fi_add_biomarker(...) %>%
  fi_calc_age(...) %>%
  fi_calc_duration(...) %>%
  fi_cohort_include(...) %>%
  fi_cohort_include(...) %>%
  fi_cohort_include(...)
```

Results

...



FlatironKitchen Functions



Our final “Recipe”

```
1 fk <- fi_cohort_start("SkinCancer") %>%
2   fi_add_demographics(...) %>%
3   fi_add_mortality(...) %>%
4   fi_add_lastcontact(...) %>%
5   fi_add_biomarker(...) %>%
6   fi_add_biomarker(...) %>%
7   fi_calc_age(...) %>%
8   fi_calc_duration(...) %>%
9   fi_cohort_include(...) %>%
10  fi_cohort_include(...) %>%
11  fi_cohort_include(...)
```



Line by line console feedback

--- `fi_cohort_start()` -----

Touching: **DIAGNOSIS**

NewCols : PatientID [c], Diagnosis [c],
DiagnosisDate [d]

Cohort : **1,000**

--- `fi_add_demographics()` -----

Touching: **DEMOGRAPHICS**

NewCols : Sex [c], BirthYear [n] ...

--- `fi_add_mortality()` -----

Touching: **MORTALITY**

Note : Imputing death date to month middle

Note : Death = !is.na(DateOfDeath)

NewCols : DateOfDeath [d], Death [l]

FlatironKitchen Functions



Our final “Recipe”

```
1 fk <- fi_cohort_start("SkinCancer") %>%
2   fi_add_demographics(...) %>%
3   fi_add_mortality(...) %>%
4   fi_add_lastcontact(...) %>%
5   fi_add_biomarker(...) %>%
6   fi_add_biomarker(...) %>%
7   fi_calc_age(...) %>%
8   fi_calc_duration(...) %>%
9   fi_cohort_include(...) %>%
10  fi_cohort_include(...) %>%
11  fi_cohort_include(...)
```



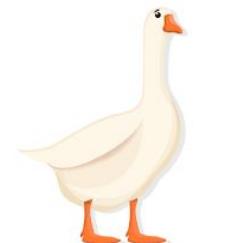
```
fk$data #Analysis-ready dataframe
```

```
# A tibble: 750 x 20
```

PatientID	Diagnosis	DiagnosisDate	Sex	BirthYear
1 A	Melanoma	2014-02-04	M	1985
2 B	Melanoma	2012-08-08	F	1976
3 C	Melanoma	2013-06-01	F	1995
4 D	Melanoma	2012-02-08	F	1984
5 E	Melanoma	2015-08-08	M	1987

... with 745 more rows, and 22 more variables:

Race <chr>, DateofDeath <date>, D
LastVisitDate <date>, LastLabDate
LastContactDate <date>, Biomarker
<chr>, BiomarkerXDiagnosisN <int>
<dbl>, DeathAge <dbl>



FlatironKitchen Functions



fi_plot_x()

fi_plot_attrition(),
fi_plot_continuous(), ...

Visualise cohort data

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983		32			α +++
B	1991		22			β -
C	1985		25			α ++
D	1979		37			β --

"Visualise cohort attrition from initial to final cohort"

```
fk %>% fi_plot_attrition()
```



All Skin Cancer Patients 1,000 (100%)

Diagnosed after 2012 900 (90%)

Positive Follow Up Time 850 (85%)

BM X Tested Before Diagnosis 750 (75%)
0

Patients (N)

Nathaniel's 29-01-2020 Skin Cancer Analysis
Patients = 750, FlatironKitchen v. 0.2.4, System Date = 2020-01-29

FlatironKitchen Functions



fi_plot_x()

fi_plot_attrition(),
fi_plot_continuous(), ...

Visualise cohort data

"Visualise cohort attrition from initial to final cohort"

```
fk %>% fi_plot_attrition()
```

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outc
A	1983	📅	32	DNA	💊 α	++
B	1991	📅	22	DNA	💉 β	-
C	1985	📅	25	DNA	💊 α	++
D	1979	📅	37	DNA	💉 β	--



All Skin Cancer Patients

1,000 (100%)

Diagnosed after 2012

900 (90%)

Nathaniel's 29-01-2020 Skin Cancer Analysis

Patients = 750, FlatironKitchen v. 0.2.4, System Date = 2020-01-29

0

Patients (N)

Note: All code and outputs are for illustration purposes only, may have been edited for clarity and confidentiality, and do not accurately reflect real patient data.

FlatironKitchen Functions



fi_plot_x()

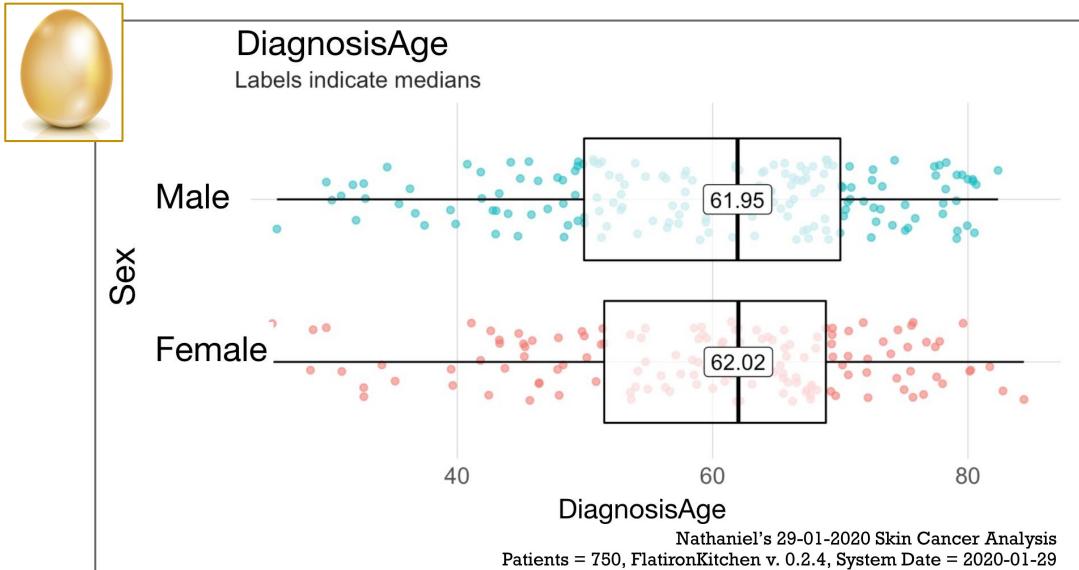
fi_plot_attrition(),
fi_plot_continuous(), ...

Visualise cohort data

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983	📅	32	DNA	💊 α	+++
B	1991	📅	22	DNA	💉 β	-
C	1985	📅	25	DNA	💊 α	++
D	1979	📅	37	DNA	💉 β	--

“Visualise age at diagnosis by sex”

```
fk %>% fi_plot_continuous("DiagnosisAge",  
                           strata = "Sex")
```



Note: All code and outputs are for illustration purposes only, may have been edited for clarity and confidentiality, and do not accurately reflect real patient data.

FlatironKitchen Functions



fi_plot_x()

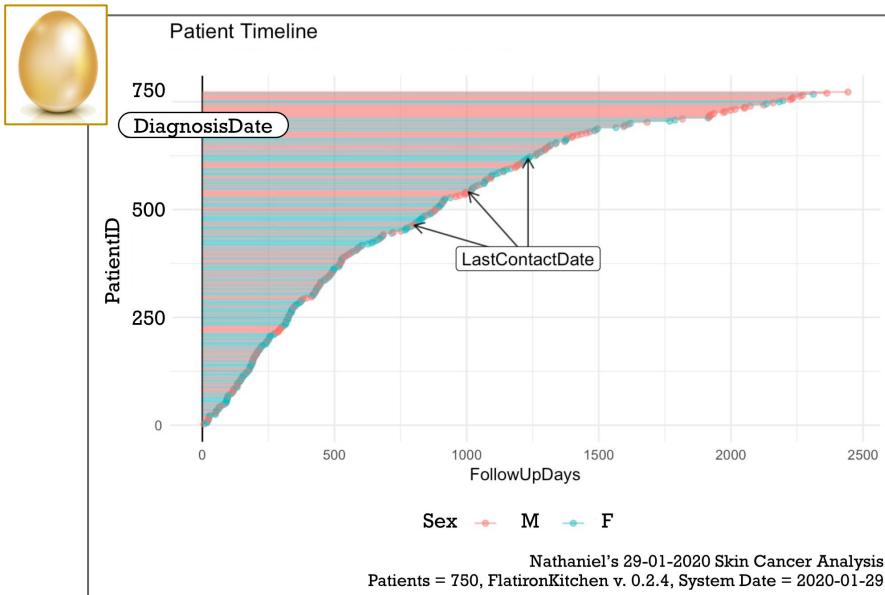
fi_plot_attrition(),
fi_plot_continuous(), ...

Visualise cohort data

PatientID	BirthYear	DxDate	DxAge	BMXDX	TxDx	Outcome
A	1983		32			α +++
B	1991		22			β -
C	1985		25			α ++
D	1979		37			β --

Create patient timelines colored by Sex

```
fk %>% fi_plot_timeline(index =  
  "DiagnosisDate", strata = "Sex")
```



Note: All code and outputs are for illustration purposes only, may have been edited for clarity and confidentiality, and do not accurately reflect real patient data.

FlatironKitchen Ecosystem



FlatironKitchen

0.2.3.9005

Functions

News

Workflows ▾

Articles ▾

Lessons ▾

Issues

Cheatsheets!

The image shows three FlatironKitchen cheatsheets:

- FlatironKitchen | FlatironKitchen Objects Cheatsheet**: A comprehensive guide to creating FlatironKitchen objects, including sections on DataMart, Project and Results, and various object types like DataFrames, DataTables, and DataSets.
- FlatironKitchen | Cohort Enhancement Cheatsheet**: A guide to enhancing cohorts, showing how to start cohorts, add filters, calculate metrics, and perform cohort analysis.
- FlatironKitchen | FlatironKitchen Workflow Cheatsheet**: A guide to workflow management, showing how to create workflows, add steps, calculate metrics, and manage results.

End-to-End Vignettes

Metastatic RCC Recipe

Jinjoo Shim (jinjoo.shim@roche.com)

Go to Recipes main page

Goal

For patients with a metastatic renal cell carcinoma diagnosis between Jun-1-2012 and Mar-31-2018, compare survival after metastatic diagnosis by GroupStage at diagnosis.

FlatironKitchen Code

```
library(FlatironKitchen)  
fk <- fi_start("MetastaticRCC") %>%  
  
# Add Demographics  
fk_add_demographics() %>%  
  
# Add Mortality  
fk_add_mortality()
```

Guided Lessons

Lesson 1

Creating and understanding FlatironKitchen objects



Go to Lessons main page

Goals

- Work with the metastatic breast cancer datamart
- Create a FlatironKitchen object
- Understand what is in a FlatironKitchen object
- Create summary outputs from a FlatironKitchen object

Tasks

FlatironKitchen Impact



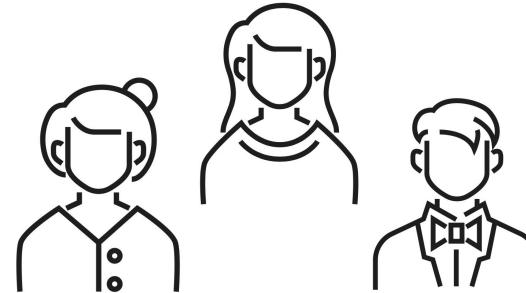
1. Consistent, elegant code and outputs.

2. Reduced time from idea to insight.

3. Increased outreach and enabling.



Epidemiologist

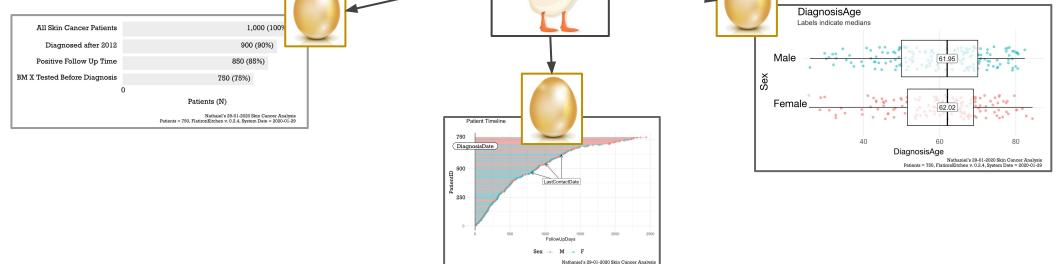


Data Scientists

```
1 library("FlatironKitchen")
2
3 fk <- fi_start("SkinCancer") %>%
4   fi_add_biomarker(...) %>%
5   fi_add_biomarker(...) %>%
6   fi_calc_age(...)
```



Biostatistician



Note: All code and outputs are for illustration purposes only, may have been edited for clarity and confidentiality, and do not accurately reflect real patient data.



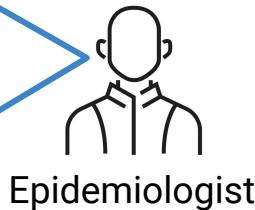
FlatironKitchen Impact

"I have always wanted to touch data but never had the time to learn how [...]

FlatironKitchen lets me finally do that [...]

It's like heaven.

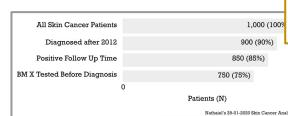
~ Genentech Epidemiologist



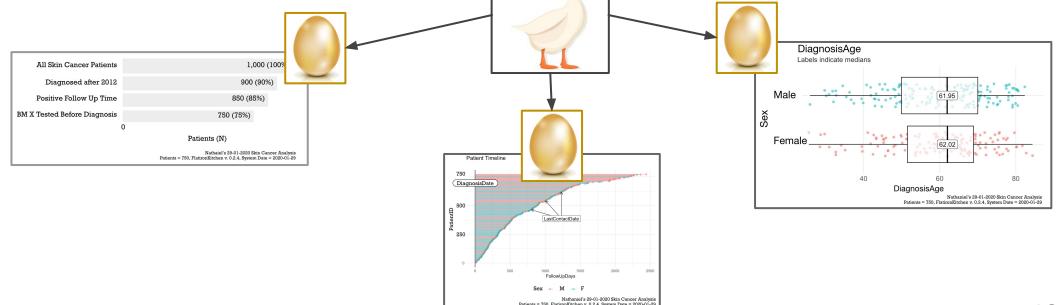
Data Scientists



Biostatistician



```
1 library("FlatironKitchen")
2
3 fk <- fi_start("SkinCancer") %>%
4   fi_add_biomarker(...) %>%
5   fi_add_biomarker(...) %>%
6   fi_calc_age(...)
```



Note: All code and outputs are for illustration purposes only, may have been edited for clarity and confidentiality, and do not accurately reflect real patient data.

FlatironKitchen Impact

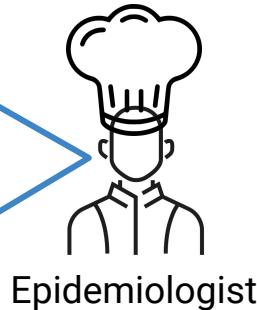


"I have always wanted to touch data but never had the time to learn how [...]

FlatironKitchen lets me finally do that [...]

It's like heaven.

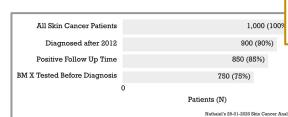
~ Genentech Epidemiologist



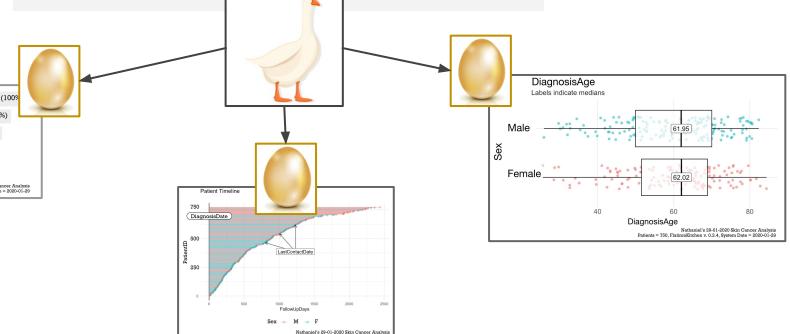
Data Scientists



Biostatistician



```
1 library("FlatironKitchen")
2
3 fk <- fi_start("SkinCancer") %>%
4   fi_add_biomarker(...) %>%
5   fi_add_biomarker(...) %>%
6   fi_calc_age(...)
```



Note: All code and outputs are for illustration purposes only, may have been edited for clarity and confidentiality, and do not accurately reflect real patient data.

FlatironKitchen

How we overhauled a Frankensteinian workflow with the tidyverse to enable fast, reproducible, elegant analyses of electronic health records.

rstudio::conf 2020, San Francisco

Nathaniel D. Phillips

On behalf of Roche / Genentech PHC Data Science Analytics

✉ phc.datascience@roche.com