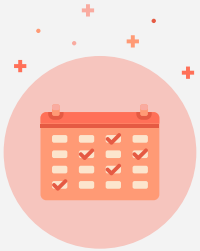
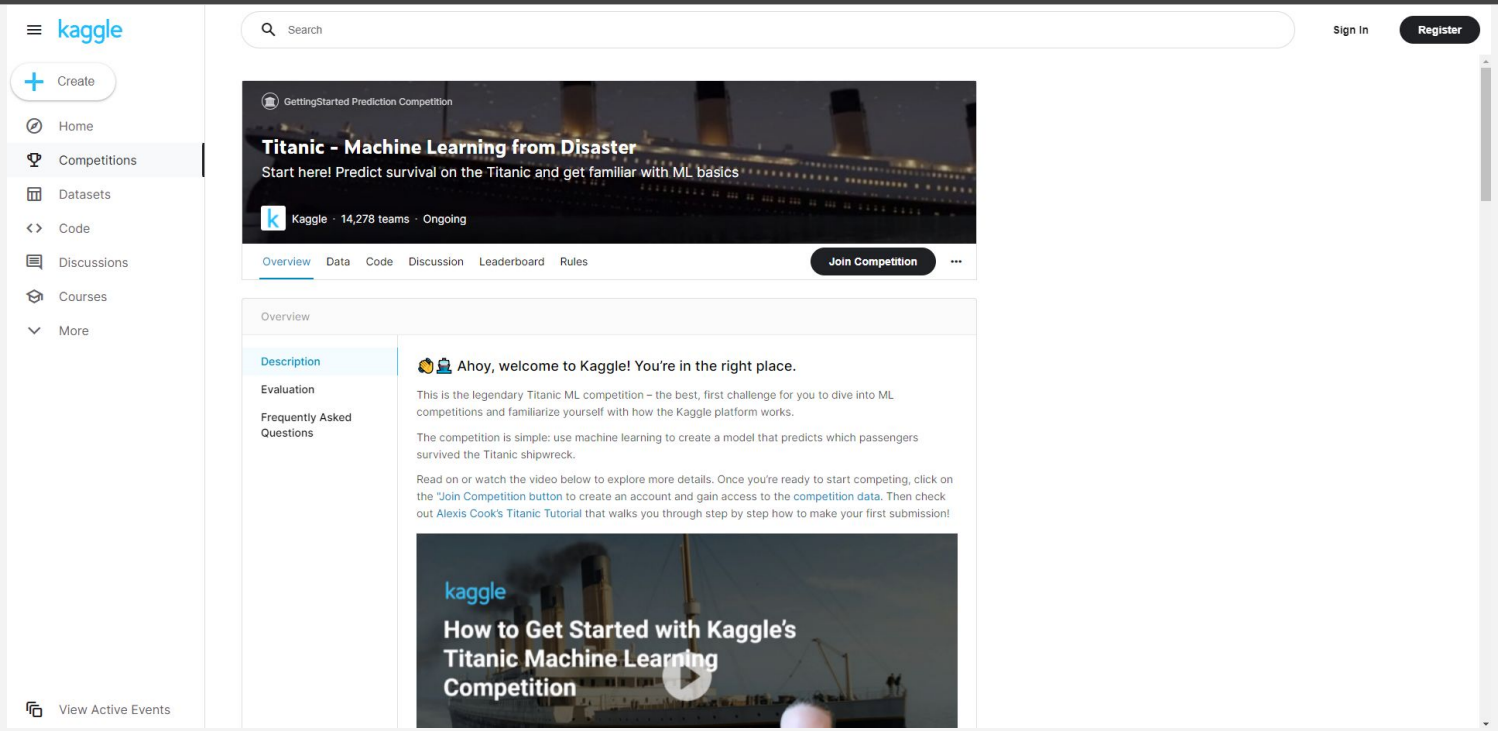


ProTips #3:
Visualization Demos

Start now!

https://www.kaggle.com/competitions/titanic/overview



The Dataset

Data Dictionary

From https://www.kaggle.com/competitions/titanic/data

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

The Dataset

```
head(titanicData, n=7)
```

A tibble: 7 × 12

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
<dbl>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<chr>	<chr>
2	Survived	1st Class	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
4	Survived	1st Class	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S
7	Passed	1st Class	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
11	Survived	3rd Class	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	16.7000	G6	S
12	Survived	1st Class	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.5500	C103	S
22	Survived	2nd Class	Beesley, Mr. Lawrence	male	34	0	0	248698	13.0000	D56	S
24	Survived	1st Class	Sloper, Mr. William Thompson	male	28	0	0	113788	35.5000	A6	S

Visualization Demos Overview

Fare~Age

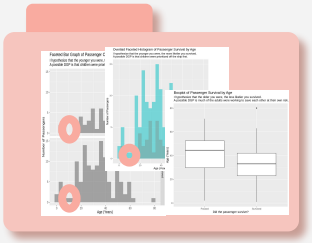
Numerical~Numerical

- Jitter Plot

Survived~Pclass

Categorical~Categorical

- Faceted Bar Graph + Grouped



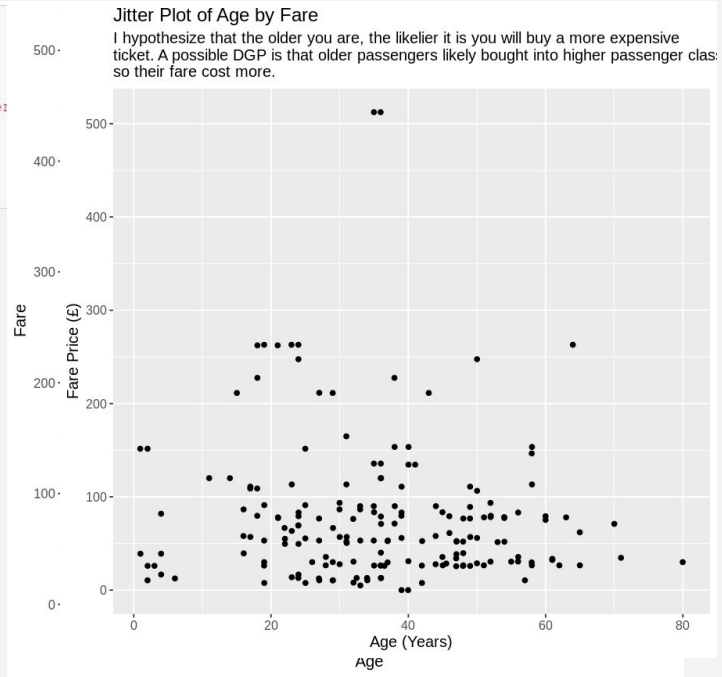
Survived~Age

Categorical~Numerical

- Faceted Histogram + Overlaid
- Boxplot

Numerical~Numerical - Jitter Plot

```
# Finalized Jitter Plot (Scatterplot is similar, but we encourage students to use Jitter)
gf_jitter(Fare~Age, data=titanicData) %>%
gf_theme(text=element_text(size=12)) +
labs(title="Jitter Plot of Age by Fare",
    subtitle="I hypothesize that the older you are, the likelier it is you will buy a more exper
ticket. A possible DGP is that older passengers likely bought into higher passenger classes,
so their fare cost more.",
    x="Age (Years)",
    y="Fare Price (£)")
```

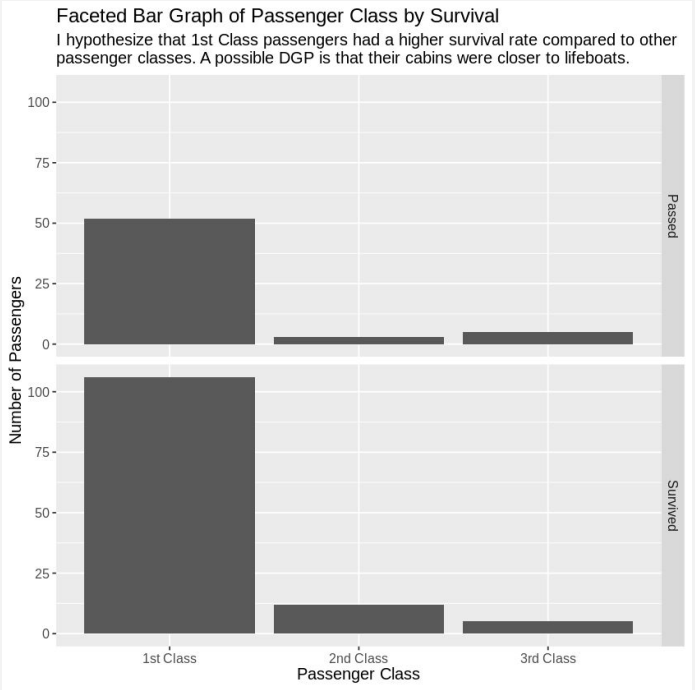


What graphs stuff:

- Title a descriptive title!
- Axis labels!
- Axis measurements!

Categorical~Categorical - Faceted Bar Graph

```
# Finalized Faceted Bar Graph
gf_bar(~Pclass, data=titanicData) %>%
gf_facet_grid(Survived~.) %>%
gf_theme(text=element_text(size=12)) +
labs(title="Faceted Bar Graph of Passenger Class by Survival",
    subtitle="I hypothesize that 1st Class passengers had a higher survival rate compared to other
passenger classes. A possible DGP is that their cabins were closer to lifeboats.",
    x="Passenger Class",
    y="Number of Passengers")
```

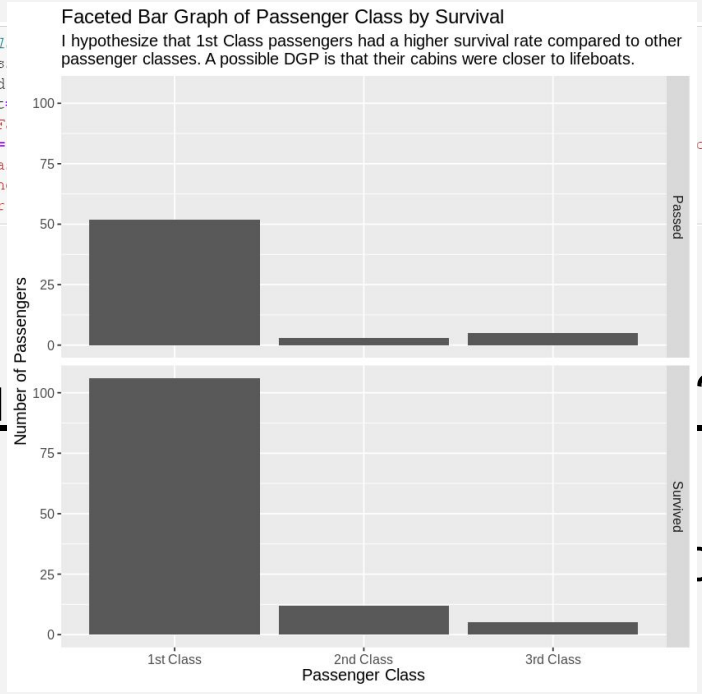


Hold Up! What's missing?

- Title a descriptive title!
- Axis labels!
- Axis measurements!

Categorical~Categorical - Colored Faceted Bar Graph

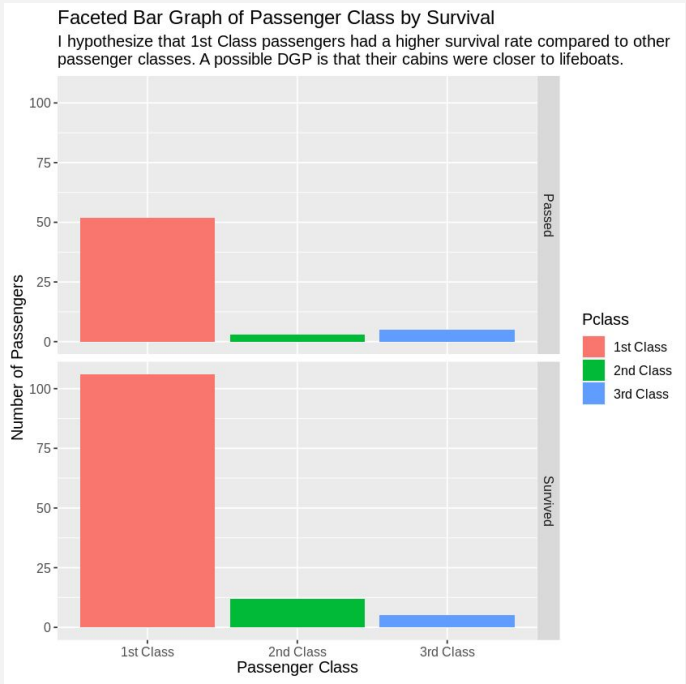
```
# Colored Pcl
gf_bar(~Pclas
gf_facet_grid
gf_theme(text
labs(title="F
    subtitle=
passenger cla
    x="Passen
    y="Number
```



ompared to other

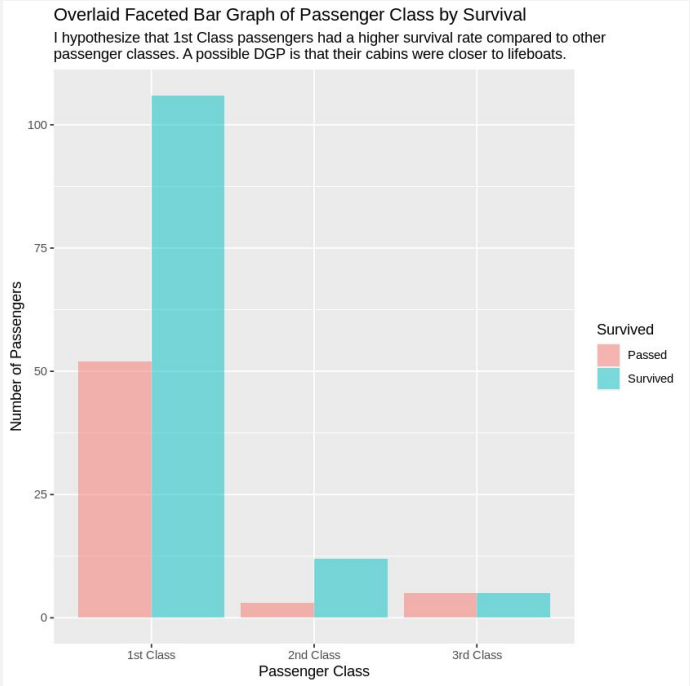
Qu ?

ch!



Categorical~Categorical - Grouped Faceted Bar Graph

```
# Finalized Overlaid Faceted Bar Graph
gf_bar(~Pclass, fill=~Survived, alpha=0.5,
       position="dodge", data=titanicData) +
gf_facet_grid(Survived~.) %>%
gf_theme(text=element_text(size=12)) +
labs(title="Overlaid Faceted Bar Graph of Passenger Class by Survival",
    subtitle="I hypothesize that 1st Class passengers had a higher survival rate compared to other
passenger classes. A possible DGP is that their cabins were closer to lifeboats.",
    x="Passenger Class",
    y="Number of Passengers")
```



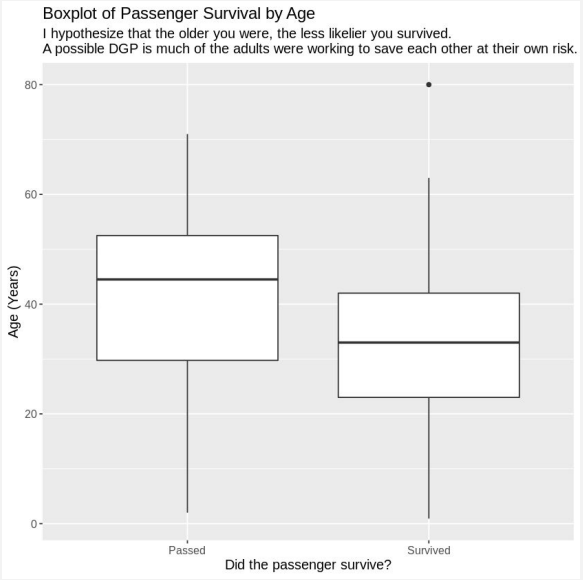
When Color Is Cool:

Overlaid gf_facet_grid()

- position="dodge"
- "+" vs. "%>%"

Categorical~Numerical - Boxplot

```
# Boxplot
gf_boxplot(Age~Survived, data=titanicData) %>%
gf_theme(text=element_text(size=12)) +
labs(title="Boxplot of Passenger Survival by Age",
    subtitle="I hypothesize that the older you were, the less likelier you survived.
A possible DGP is much of the adults were working to save each other at their own risk.",
    x="Did the passenger survive?",
    y="Age (Years)")
```

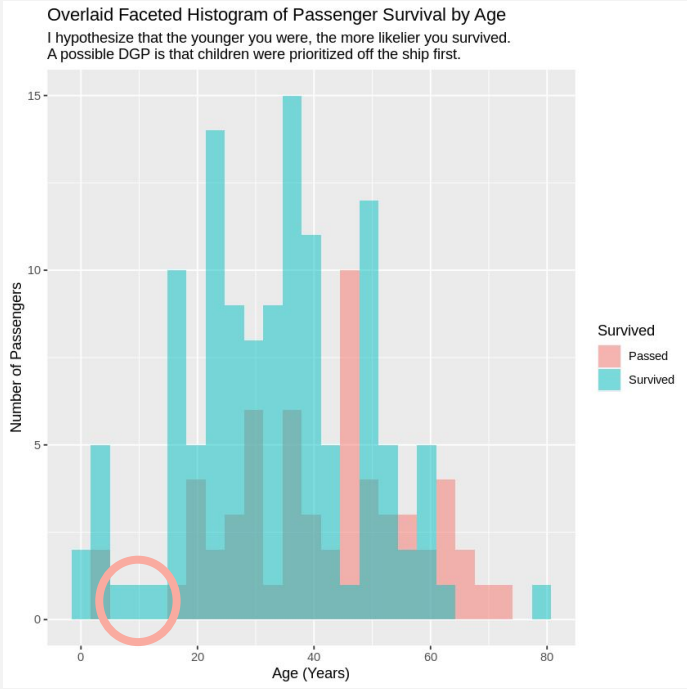


Categorical~Numerical - Overlaid Faceted Histogram Graph

```
# Overlaid Faceted Histogram
gf_histogram(~Age, fill=~Survived, alpha=0.5,
             position="identity", data=titanicData) +
gf_facet_grid(Survived~.) %>%
gf_theme(text=element_text(size=12)) +
labs(title="Overlaid Faceted Histogram of Passenger Survival by Age",
    subtitle="I hypothesize that the younger you were, the more likelier you survived.
A possible DGP is that children were prioritized off the ship first.",
    x="Age (Years)",
    y="Number of Passengers")
```

Overlaid gf_facet_grid()

- position="identity"



No really, any questions?

Any questions? :')

A Reminder About Office Hours

Teaching Team

Name	Role	Section ID	Section Time	Office Hours
Judith Fan	Instructor	N/A	N/A	Thurs 11am-12pm in McGill Hall 5141
Simran Barnwal	TA	75428 (A05)	Wed 2pm	Wed 3:30pm-4:30pm in Zoom
Amy Fox	TA	75424 (A01), 75425 (A02)	Tues 11am, Tues 12pm	TBD in Zoom
Holly Huey	TA	75429 (A06), 75430 (A07)	Thurs 5pm , Fri 10am	Thurs 11am-12pm in Zoom
Zoe Tait	TA	75424 (A01)	Tues 11am	Thurs 11am-12PM in Zoom
Vryan Feliciano	TA	75430 (A07)	Fri 10am	Wed 4pm-5pm at MOM Cafe and Zoom
Lea Bronnimann	TA	75427 (A04)	Wed 10am	Wed 11am-12pm in Zoom
Justin Yang	TA	75426 (A03)	Tues 1pm	Tues 2pm-3pm in Zoom